Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared to models based on simple clinical features: an evaluation using clinical trial data

Authors: John M Dennis, Beverley M Shields, William E Henley, Angus G Jones, Andrew T Hattersley*

Author details:

1. John M Dennis (MSc). University of Exeter Medical School. Address: Institute of Biomedical & Clinical Science, RILD Building, Royal Devon & Exeter Hospital, Barrack Road, Exeter EX2 5DW, UK
2. Beverley M Shields (PhD). University of Exeter Medical School. Address: Institute of Biomedical & Clinical Science, RILD Building, Royal Devon & Exeter Hospital, Barrack Road, Exeter EX2 5DW, UK
3. William E Henley (PhD). University of Exeter Medical School. Address: Health Statistics Group, Institute of Health Research, University of Exeter Medical School, Exeter EX1 2LU, UK
4. Angus G Jones (PhD). University of Exeter Medical School. Royal and Devon Exeter Hospital. Address: University of Exeter Medical School. Address: Institute of Biomedical & Clinical Science, RILD Building, Royal Devon & Exeter Hospital, Barrack Road, Exeter EX2 5DW, UK
5. Andrew T Hattersley (DM). University of Exeter Medical School. Address: Institute of Biomedical & Clinical Science, RILD Building, Royal Devon & Exeter Hospital, Barrack Road, Exeter EX2 5DW, UK

*Correspondence to:
Andrew Hattersley: Email A.T.Hattersley@exeter.ac.uk, phone +44 1392 408260

Abstract

Background

Recent research using data-driven cluster analysis has proposed five subgroups of diabetes with differences in diabetes progression and risk of complications. We aimed to compare the clinical utility of this subgroup-based approach for predicting patient outcomes with an alternative strategy of developing models for each outcome using simple patient characteristics.

Methods

We identified clusters in the ADOPT (n=4,351) trial cohort using the cluster analysis reported by Ahlqvist and colleagues (Lancet Diabetes Endocrinology 2018;6:361-69). Differences between clusters in glycaemic and renal progression were evaluated, and contrasted with stratification using simple continuous clinical features (respectively, age at diagnosis and baseline renal function). We tested the performance of a strategy of selecting glucose-lowering therapy using clusters with one combining simple clinical features (sex, BMI, age at diagnosis, baseline HbA1c) in an independent trial (RECORD (n=4,447)).

Findings

Clusters identified in trial data were similar to those described in the original study. Clusters showed differences in glycaemic progression, but a model with age at diagnosis alone explained a similar amount of variation in progression. We found differences in CKD incidence between clusters however baseline eGFR was a better predictor of time to CKD. Clusters differed in glycaemic response, with a particular benefit for cluster 3 (insulin-resistant) with

thiazolidinediones and cluster 5 (older) with sulfonylureas. However simple clinical features outperformed clusters to select therapy for individual patients.

Interpretation

The proposed data-driven clusters differ in diabetes progression and treatment response, but models based on simple continuous clinical features are more useful to stratify patients. This suggests precision medicine in type 2 diabetes is likely to have most clinical utility if based on an approach of using specific phenotypic measures to predict specific outcomes, rather than assigning patients into subgroups.

Funding

Medical Research Council (UK) (MR/N00633X/1).

Research in context

Evidence before this study

A recent study published in *Lancet Diabetes and Endocrinology* proposed a novel substratification of diabetes, using a data-driven cluster analysis in Scandinavian registry data to identify five reproducible subgroups of adult-onset diabetes. The authors went on to show differences between the clusters in disease progression and risk of complications in observational follow-up. The authors suggested the clusters might help with therapy selection in the future but did not test whether the clusters could inform therapy choice. We searched Scopus, Web of Science, and Google Scholar to track the citations of the original study, searching for follow-up studies assessing the reproducibility, clinical utility and role in treatment selection of the proposed data-driven clusters up to January 1, 2019. We identified a study that identified similar clusters in Chinese and a small mixed American population but did not examine any aspect of clinical utility as clinical follow-up was not available. A second study of Danish patients applied similar cluster analysis and, with duration of diabetes as an additional input variable, identified five subgroups of type 2 diabetes that differed to those in the original study, and differed in the prevalence of diabetes complications. No studies were found that tested the clinical utility and particularly the role in treatment of the proposed cluster-based approach.

Added value of this study

This study takes forward the concept of heterogeneity in type 2 diabetes, by testing the clinical utility of the data-driven cluster approach proposed by

Ahlqvist and colleagues. The cluster analysis was repeated, and differences by cluster in disease progression and treatment response were evaluated in newly diagnosed participants in the ADOPT trial with randomised, protocol-driven follow-up data available. We found the clusters were reproducible and did differ in progression and treatment response. However, simpler clinical measures were as or more useful than the clusters for stratifying each outcome assessed.

Implications of all the available evidence

Patients with type 2 diabetes differ in treatment response and risk of disease progression, raising the possibility of a practical, clinically orientated stratified approach in the near future. Our study suggests a ‘prediction model’ approach, combining phenotypic measures to predict specific outcomes for individual patients, is likely to have greater clinical utility than approaches that use clinical features to assign individuals into subgroups.

Introduction

Type 2 diabetes is a heterogeneous multifactorial condition, comprising 90-95% of all diabetes and affecting over 400 million people worldwide. There is currently great interest in better characterising the heterogeneity in type 2 diabetes, and in exploiting this heterogeneity to improve care and outcomes for individuals with type 2 diabetes.(1-3)

In a recent study, Ahlqvist and colleagues identified five replicable clusters of individuals with diabetes in Scandinavian registry data.(4) The smallest cluster was defined by the presence of glutamic acid decarboxylase (GAD) autoantibody positivity, regardless of other characteristics (Cluster 1: severe autoimmune diabetes (SAID)). Four 'type 2' like clusters were then characterised by the absence of GAD positivity and varying degrees of differences in age at diagnosis, and baseline measures of BMI, HbA1c, and HOMA2 measured insulin resistance and beta-cell function. The four 'type 2' clusters were named as follows; Cluster 2: severe insulin deficient diabetes (SIDD); Cluster 3: severe insulin resistant diabetes (SIRD); Cluster 4: mild obesity-related diabetes (MOD); Cluster 5: mild age-related diabetes (MARD). Ahlqvist and colleagues then showed potentially clinically important differences in disease progression and risk of complications between the clusters in observational follow-up, most notably a striking increase in the risk of diabetic kidney disease in the cluster characterised by insulin resistance (Cluster 3: SIRD).

The key question for any subgroup analysis is its clinical utility, and in particular whether the proposed subgroups differ in response to therapy which could help inform treatment strategies.(2) Ahlqvist and colleagues suggested but did not demonstrate that the clusters could be useful to guide choice of therapy.(5) To

date the only stratified approaches in type 2 diabetes showing large differences in response between treatments have used subgroups defined by routine clinical measures such as sex and BMI.(6) A further key question, raised by van Smeden and colleagues in response to the original study, is whether assigning individuals to clusters has greater clinical utility for predicting outcomes than an approach that combines continuous clinical features to predict outcomes for individual patients.(7)

We aimed to establish the clinical utility of the clusters by analysing two large existing trial datasets of individuals randomised to metformin, sulfonylurea and thiazolidinediones therapy, ADOPT and RECORD.(8, 9) In contrast to the observational follow-up in the original study of Scandinavian registry data, these trial datasets provided protocol-driven, randomised follow-up to evaluate clinical outcomes and differences in response to therapy. We compared the utility of the data-driven clusters with simpler approaches based on routine clinical measures available in any diabetes clinic.

Methods

Study population

The primary study population comprised newly diagnosed, drug-naïve, individuals with type 2 diabetes participating in the ADOPT trial of glycaemic durability, randomised to metformin, sulfonylurea (glibenclamide) or thiazolidinedione (rosiglitazone) monotherapy up to five years (n=4,351).(8) Eligibility criteria at screening included: Age 30-75 years, fasting plasma glucose 7-13 mmol/l, no evidence of renal impairment (serum creatinine >114 μmol/l for males or >106 μmol/l for females). As a replication dataset we used the RECORD study (n=4,447), a cardiovascular outcomes trial in individuals with established type 2 diabetes (mean duration of diabetes 7 years) initiating

the same drug classes as ADOPT but as dual second-line therapy, for up to six-years.(9) Sulfonylurea type was based on local practice (glibenclamide[18%], gliclazide[30%], or glimepiride[52%]), rosiglitazone was the thiazolidinedione used. Eligibility criteria included: Age 40-75 years, BMI >25.0 kg/m², HbA1c >7.0% and ≤9.0%, and no evidence of renal impairment (serum creatinine >130 µmol/l).

We followed individuals in both trials from randomisation until the earliest of: the primary outcome of the original trial; censor date, five years, or the occurrence of an outcome of interest. Full individual level trial data were accessed through Clinical Trial Data Transparency Portal (Proposal 930).

Measurements

We calculated HOMA2 measures of insulin resistance and beta-cell function using fasting C-peptide and fasting-glucose measures using the HOMA 2 calculator.(10) In ADOPT GAD antibody positivity (yes or no) was measured using a commercially available radioimmunoassay.(11) In RECORD all required measures except GAD were measured at baseline, we calculated HOMA2 measures using fasting insulin as fasting C-peptide was not available. Sex, age at diagnosis, baseline BMI and baseline HbA1c comprised the other measures required for cluster analysis.

Definitions of study outcomes

Glycaemic progression

Glycaemic progression was defined as the change in HbA1c from one year up to five years (HbA1c at time *t* – HbA1c at one year), thus allowing for an initial period of treatment response up to one year.

Kidney disease

Chronic kidney disease (CKD) was defined as progression from normal GFR (eGFR ≥60 ml/min per 1.73m²) to confirmed CKD Stage 3 (two consecutive measures of eGFR <60 ml/min per 1.73m²). eGFR was calculated using CKD-EPI; as a sensitivity analysis eGFR was also calculated using MDRD.(12) Measures of renal function were recorded at baseline, six months and annually. If progression was confirmed, the first of the two study visits was used to define CKD onset. Albuminuria was defined as progression from normal urinary albumin to creatinine ratio (UACR) (UACR <30 mg/g) to either microalbuminuria (UACR 30-300 mg/g) or macroalbuminuria (UACR ≥300 mg/g). Individuals with eGFR <60 and UACR ≥30 at their baseline visit were excluded from, respectively, the analysis of CKD and albuminuria outcomes.

Glycaemic response

HbA1c was evaluated as achieved HbA1c and as cumulative HbA1c reduction at three years as measured by area-under-the-curve (3 year AUC HbA1c). AUC HbA1c is equivalent to the time-updated HbA1c measure used in the UK Prospective Diabetes Study outcomes model.(13) Three years was chosen as the time point at which average AUC HbA1c was approximately equal between the three drugs.(8) Other time points will tend to favour a specific therapy; early time points will favour sulfonylureas as these agents have greater short-term response, whilst later time points favour thiazolidinediones which have greater glycaemic durability.(8)

Statistical analysis

Cluster analysis

In ADOPT, we repeated the clustering approach of Ahlqvist and colleagues.(4) Males and females were clustered separately then pooled, continuous

measures were mean centred and standardised, and continuous measures >5 standard deviations from the mean were excluded. K-means clustering specifying four clusters was applied to the GAD-negative subset of individuals as K-means clustering does not incorporate binary variables; all GAD-positive individuals were manually assigned to a separate cluster.(4) The same R command (kmeansrun), number of runs (100) and measure of cluster stability (Jaccard coefficient >0.75 after 2000 bootstraps) were applied.(14) Once clusters were defined we assigned the same cluster names as in the original study, based on the distribution of cluster characteristics. In RECORD, we 1) assigned each individual to their ADOPT-derived cluster based on their Euclidean distance from each cluster centre; 2) repeated the cluster analysis to derive RECORD-specific clusters. As GAD was not available, all individuals in RECORD were assumed to be GAD-negative.

Glycaemic progression

In both trials, mean HbA1c trajectories from randomisation up to five years for each cluster were first estimated using a repeated-measures mixed-effects model, including fixed effects for study visit, assigned cluster, and a study visit by cluster interaction. Patient-level random effects and an unstructured covariance matrix were specified for this and subsequent mixed-effects models. All individuals within a trial were pooled, regardless of randomised therapy. To estimate glycaemic progression by cluster the same model was then fitted but with HbA1c change from one year as the outcome. We estimated the mean annual rate of glycaemic progression for each cluster by updating the cluster model to replace study visit with time as a linear covariate. Mean HbA1c by age was estimated using the same model but a linear term for continuous age at diagnosis replacing the clusters. For each model we estimated the proportion of

variance explained (R^2) by the fixed effects, the AIC, and the adequacy index.(15, 16)

Kidney disease

We compared the cumulative incidence of CKD by cluster, using Kaplan-Meier plots and unadjusted and baseline eGFR (a continuous linear term) adjusted Cox proportional hazard models with cluster as a categorical variable. We estimated R^2 and the discrimination ability (Harrell's C-index) of the unadjusted cluster Cox model, compared with a Cox model with continuous baseline eGFR as a linear term.(16) We repeated the same analysis for time to a 30% decline in eGFR, and for time to albuminuria with and without adjustment for baseline UACR as a continuous linear term. We also compared continuous relative changes from baseline in eGFR and UACR progression over 0-5 years by cluster using a mixed-effects models with fixed effects for study visit, cluster, and study visit by cluster interaction.

Glycaemic response

We first evaluated whether HbA1c response to the three drugs differed across the clusters in ADOPT. Average HbA1c trajectories by drug were estimated up to three years for each cluster separately, using repeated-measures mixed-effects models with fixed effects for study visit, drug, visit by drug interaction and visit by baseline HbA1c interaction. 3 year AUC HbA1c was estimated for each drug in each cluster as the integral of the area under the mean HbA1c trajectory using the trapezoidal rule.

Treatment selection based on HbA1c – are clusters or clinical features more useful to guide therapy?

We evaluated whether clusters were more useful than simple clinical features to select a drug for individual patients based on predicted 3 year AUC HbA1c.

Models to predict HbA1c were developed in ADOPT using two strategies: A) using the clusters and B) using clinical features. For the clusters strategy we simply estimated HbA1c response for each drug at the cluster level and applied this to all individuals within the cluster. This strategy treats individuals within a cluster as homogenous for treatment response to a particular drug. For the clinical features strategy we combined sex and linear terms for age at diagnosis, baseline BMI and baseline HbA1c (the 4 routine clinical features informing the clusters) in a multivariable model to estimate HbA1c response specific to each individual for each drug. The benefit of using each strategy developed in ADOPT to select treatment for individuals was then tested in an external trial population: RECORD.

1) Model development - ADOPT

Strategy A) clusters model: 3 year AUC HbA1c for each drug was estimated at cluster level as detailed in the first step (Statistical analysis: Glycaemic response). Strategy B) clinical features model: 3 year AUC HbA1c, as defined above, was estimated for each individual based on their precise clinical characteristics, using multivariable repeated-measures mixed-effects models for each drug. Each model had HbA1c up to 3 years as the outcome with age at diagnosis, BMI, baseline HbA1c and study visit by baseline HbA1c interaction as continuous linear terms, and study visit and sex as fixed effects. Model performance for each strategy was assessed using R^2.

2) Assessment of the treatment selection strategy in independent data – RECORD

The purpose of a treatment selection model is to select the most effective therapies for individual patients, and therefore improve outcome at a population level, rather than to predict drug response accurately. This means the true test

of a treatment selection model is whether it can robustly identify individuals likely to benefit from particular therapies.(17) Standard model performance metrics test the ability of a model to predict the outcome, and are therefore of limited use in this context.(17, 18)

We therefore applied the following steps to test the effectiveness of each treatment selection strategy. For each individual in RECORD, we applied the models developed in ADOPT to obtain estimates of 3 year AUC HbA1c on each drug. Under Strategy A) these predictions were according to the individual's assigned cluster (the same for all individuals within a cluster). Under Strategy B) predictions were at the individual level estimated from precise clinical features. For each strategy, we then applied a simple decision rule to assign individuals into two groups, one 'concordant' and one 'discordant'. Discordant individuals were those randomised to a drug with a predicted 3 mmol/mol higher 3 year AUC HbA1c (i.e. less improvement in HbA1c) than the drug predicted to be their best drug; all other individuals were defined as concordant.(19) The effectiveness of each treatment selection strategy was determined by the difference in 3 year AUC HbA1c between the concordant and discordant groups. 3 year AUC HbA1c by concordant/discordant group was estimated as previously described from a mixed-effects model with study visit, concordant/discordant group, baseline HbA1c, study visit by concordant/discordant group interaction and visit by baseline HbA1c interaction as fixed effects. We tested the sensitivity of results to the HbA1c threshold used to define concordance by repeating the analysis at HbA1c thresholds of 0, 1, 2 and 4 mmol/mol. All analyses were conducted using R version 3.4.1.

Cardiovascular outcomes

In RECORD we compared the time to the trial primary outcome, cardiovascular hospitalisation or cardiovascular death, by cluster using unadjusted and baseline age adjusted Cox proportional hazard models.

Assignment of clusters in ADOPT based on cluster centre coordinates from the Swedish ANDIS cohort

We assigned individuals in ADOPT to their ANDIS cluster based on their Euclidean distance from the cluster centres published by Ahlqvist and colleagues for the ANDIS cohort.(4) We then estimated glycaemic and renal progression and HbA1c response for each ANDIS-derived cluster, and compared model performance of the ADOPT defined clusters and ANDIS clusters.

Role of the funding source

The funders had no role in the study design, collection, analysis, data interpretation or writing of the report. The corresponding author had full access to all data and had final responsibility for the decision to submit for publication.

Results

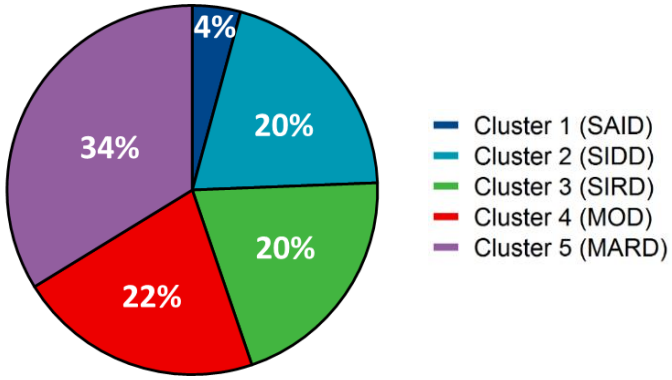
We found the clusters identified by Ahlqvist and colleagues were reproducible in trial populations. 4,003 individuals in ADOPT had valid baseline measures for cluster assignment. Of these, 3,802 were in the intention-to-treat population and so were eligible for analysis of patient outcomes. We found a clear pattern of differences between clusters in clinical characteristics (Figure 1A, Supplementary Tables 1, 3, 4), and were able to assign the same cluster names as Ahlqvist and colleagues (Figure 1B). Clusters were reasonably stable (Jaccard mean range: males 0.76-0.82; females 0.69-0.82). Cluster-centre coordinates are shown in Supplementary Table 2. In RECORD 4,148

individuals were eligible for cluster assignment (4,057 in intention-to-treat population). RECORD clusters were similar to the ADOPT clusters whether assigned from ADOPT or defined de-novo in RECORD (Supplementary Figure 1).

Average HbA1c trajectories by cluster from randomisation to five years are shown in the Supplementary Figure 2. Glycaemic progression from one year differed by cluster in ADOPT (Figure 2A), with a higher rate of progression in Clusters 1 (SAID), 2 (SIDD) and 4 (MOD). In RECORD only Cluster 4 (MOD) had a higher rate of progression (Supplementary Figure 3). However, in both trials older age at diagnosis was associated with a lower rate of glycaemic progression (mean annual difference in rate of HbA1c change per year increase in age at diagnosis: ADOPT -0.06 mmol/mol (95% confidence intervals -0.07 to -0.05; RECORD -0.05 mmol/mol (95%CI -0.06 to -0.04)) (Figure 2B, Supplementary Figure 3). Age at diagnosis explained a similar proportion of variation in progression to the clusters (ADOPT R^2=0.09 age at diagnosis, R^2=0.08 clusters; RECORD R^2=0.05 age at diagnosis, R^2=0.05 clusters). Other measures of model performance were also similar (Supplementary Table 5).

Figure 1: **Cluster distribution and cluster characteristics in ADOPT (n=4,003)**. SAID=severe autoimmune diabetes. SIDD=severe insulin-deficient diabetes. SIRD=severe insulin-resistant diabetes. MOD=mild obesity-related diabetes. MARD=mild age-related diabetes. HOMA2-B=homoeostatic model assessment 2 estimates of β-cell function. HOMA2-IR=homoeostatic model assessment 2 estimates of insulin resistance.

(A) Distribution of ADOPT participants according to k-means clustering



(B) Distributions of HbA1c, BMI, age at diagnosis, HOMA2-B, and HOMA2-IR at baseline for each cluster

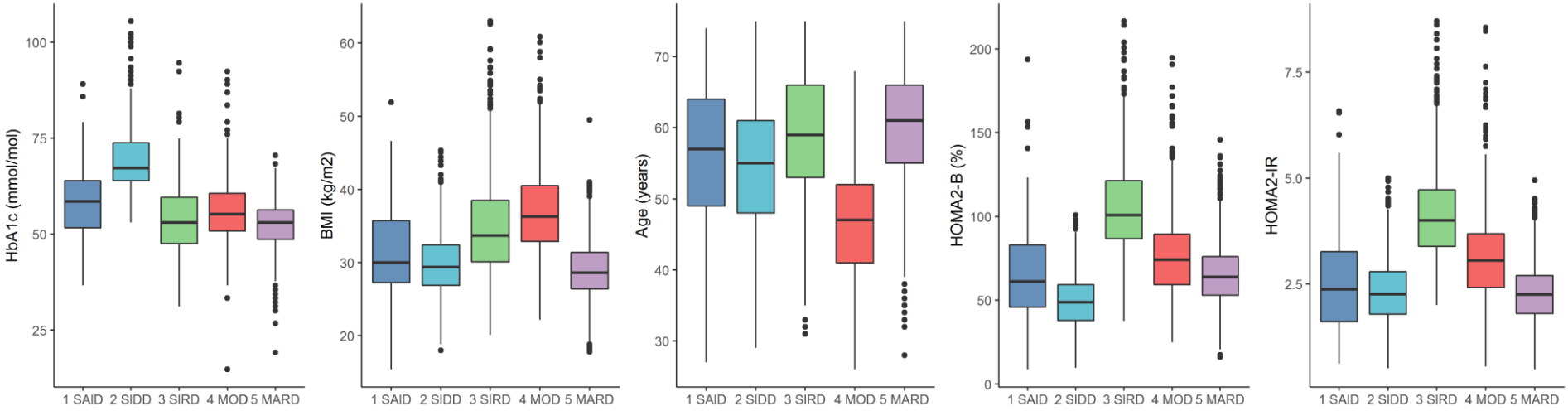
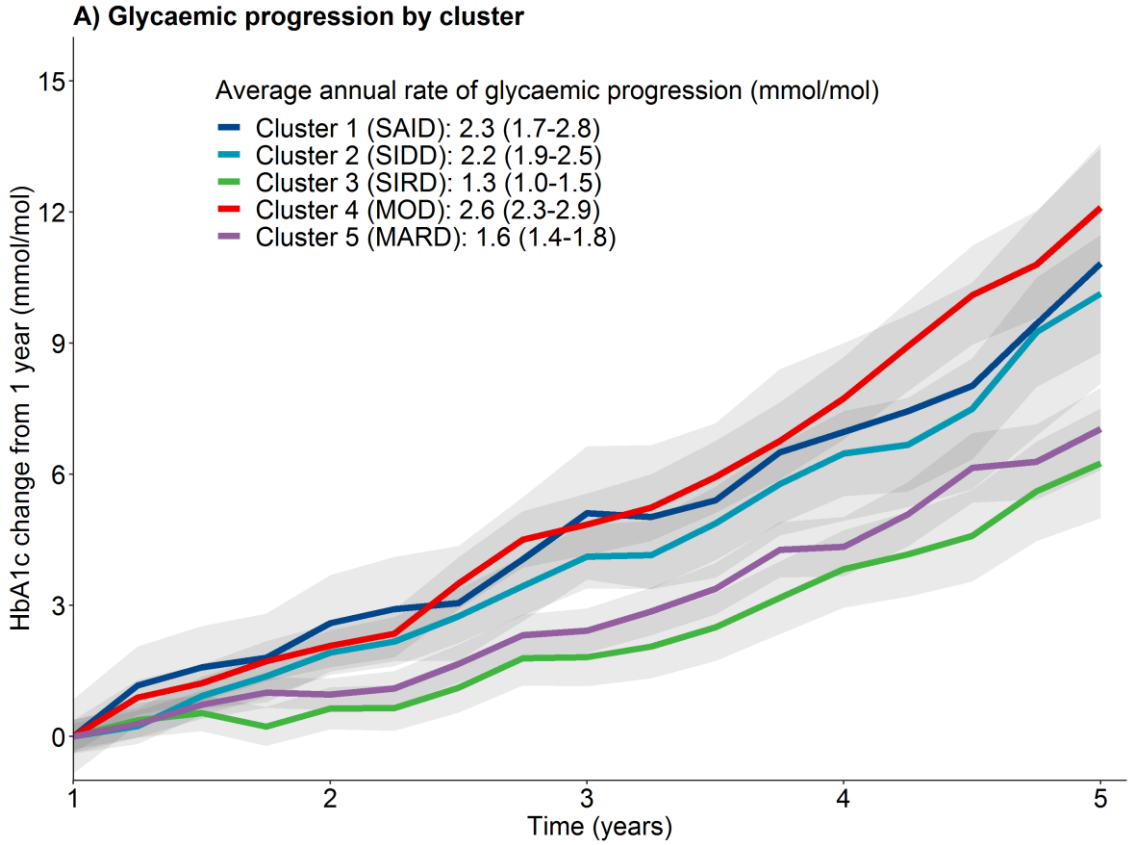
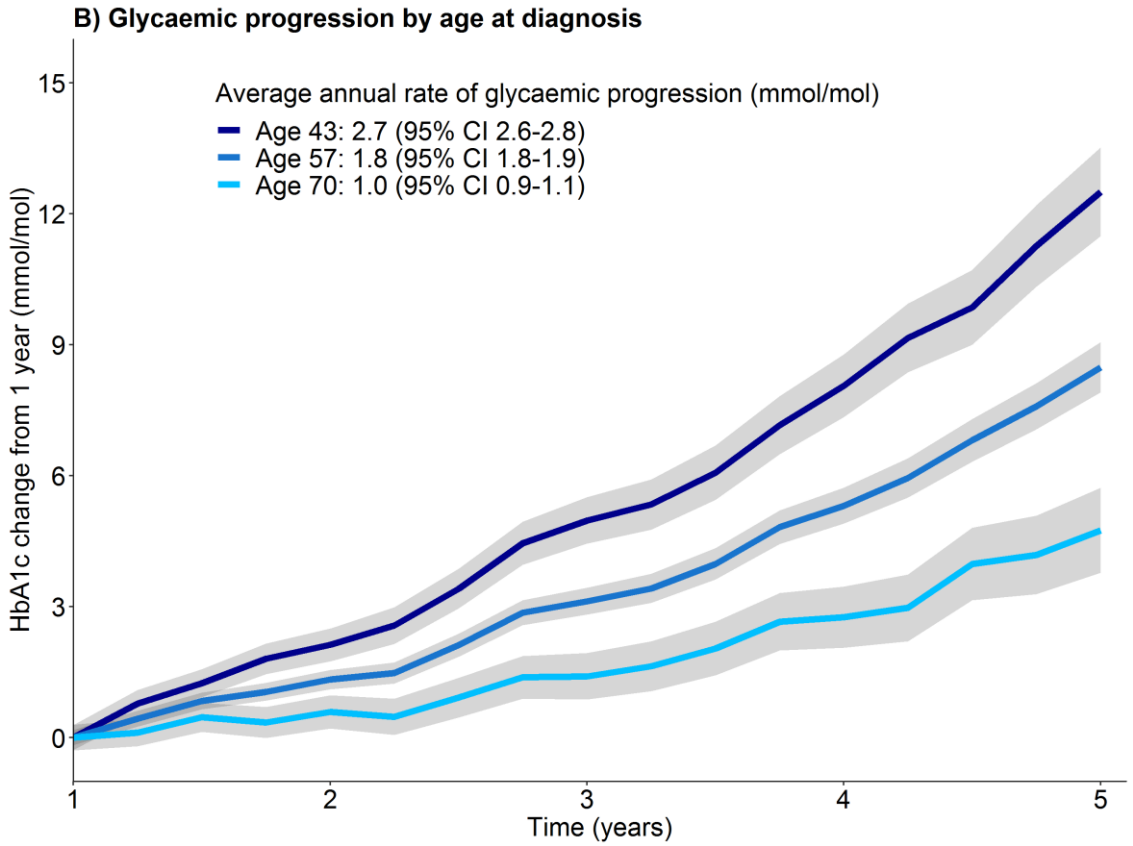


Figure 2: **Glycaemic progression by cluster in ADOPT from one to five years** A) HbA1c change by cluster (n=3,016); B) HbA1c change by age at diagnosis (10th, 50th and 90% percentile of ADOPT participants) (n=3,016). Data are estimates from repeated measures mixed-effects models.



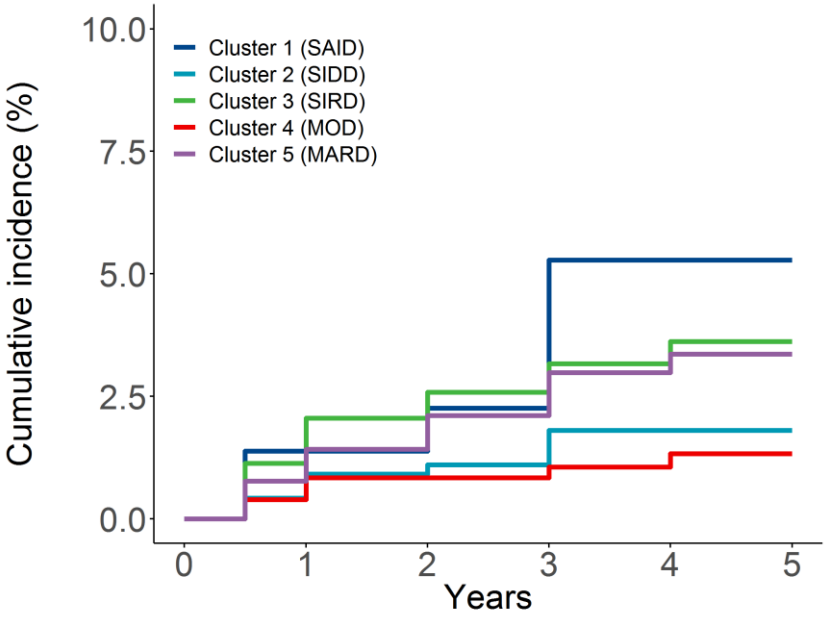


We found differences in the incidence of CKD between clusters after excluding patients with pre-existing CKD; clusters 1, 3 and 5 had the highest incidence of CKD (Figure 3A**,** Supplementary Figure 4). However, there were differences between the clusters in baseline renal function: the clusters with the highest incidence of CKD had the lowest eGFR (Supplementary Table 4). After adjustment for baseline eGFR there was no evidence of a difference in time to CKD across the clusters (Table 1, Supplementary Table 6). Results were similar using MDRD calculated eGFR (Supplementary Table 7). In ADOPT baseline eGFR explained a greater proportion of variation (R^2=0.18) and discrimination ability (C-statistic 0.90) than the clusters (R^2=0.01, C-statistic=0.58); this was similar to results in RECORD (baseline eGFR R^2=0.15, C-statistic 0.86; clusters R^2=0.01, C-statistic=0.57). Relative change from baseline in eGFR and time to 30% decline in eGFR did not differ by cluster (Supplementary Figures 5-6, Supplementary Table 8).

There was no clear pattern of difference between clusters in baseline UACR (Supplementary Table 4), in incidence of albuminuria (Figure 3B**,** Supplementary Figure 4), or in relative change in UACR (Supplementary Figure 7). After adjustment for baseline UACR time to albuminuria was shorter for cluster 3 (SIRD) versus cluster 2 (SIDD) in ADOPT, but not RECORD (Table 1, Supplementary Table 6). The clusters had no prediction and discrimination ability (ADOPT R^2=0.00, C-statistic=0.52; RECORD R^2=0.00, C-statistic=0.52), baseline UACR was a more useful measure (ADOPT R^2=0.12, C-statistic=0.74; RECORD R^2=0.10, C-statistic=0.73).

Figure 3: **Renal progression by cluster in ADOPT over five years.**

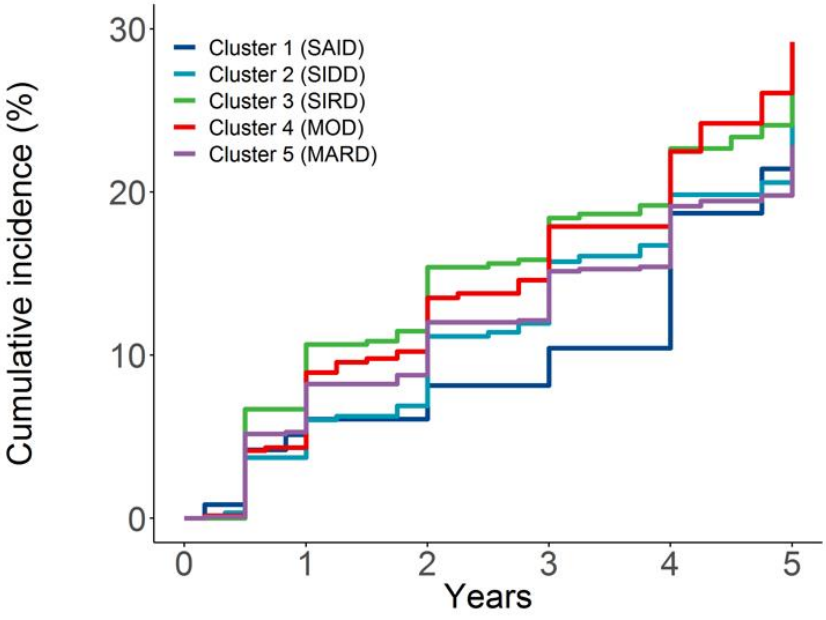
(A) Cumulative incidence of CKD Stage 3 (confirmed eGFR <60) in individuals with eGFR ≥60 at baseline (n=3,694)



Number at risk

Cluster 1 (SAID)	155	127	112	97	80	38
Cluster 2 (SIDD)	758	614	522	423	344	138
Cluster 3 (SIRD)	743	644	559	499	428	153
Cluster 4 (MOD)	808	674	559	450	366	125
Cluster 5 (MARD)	1230	1080	995	897	769	297
	0	1	2	3	4	5

(B) Cumulative incidence of albuminuria (UACR ≥30 mg/g) in individuals with UACR <30 mg/g at baseline (n=3,168).

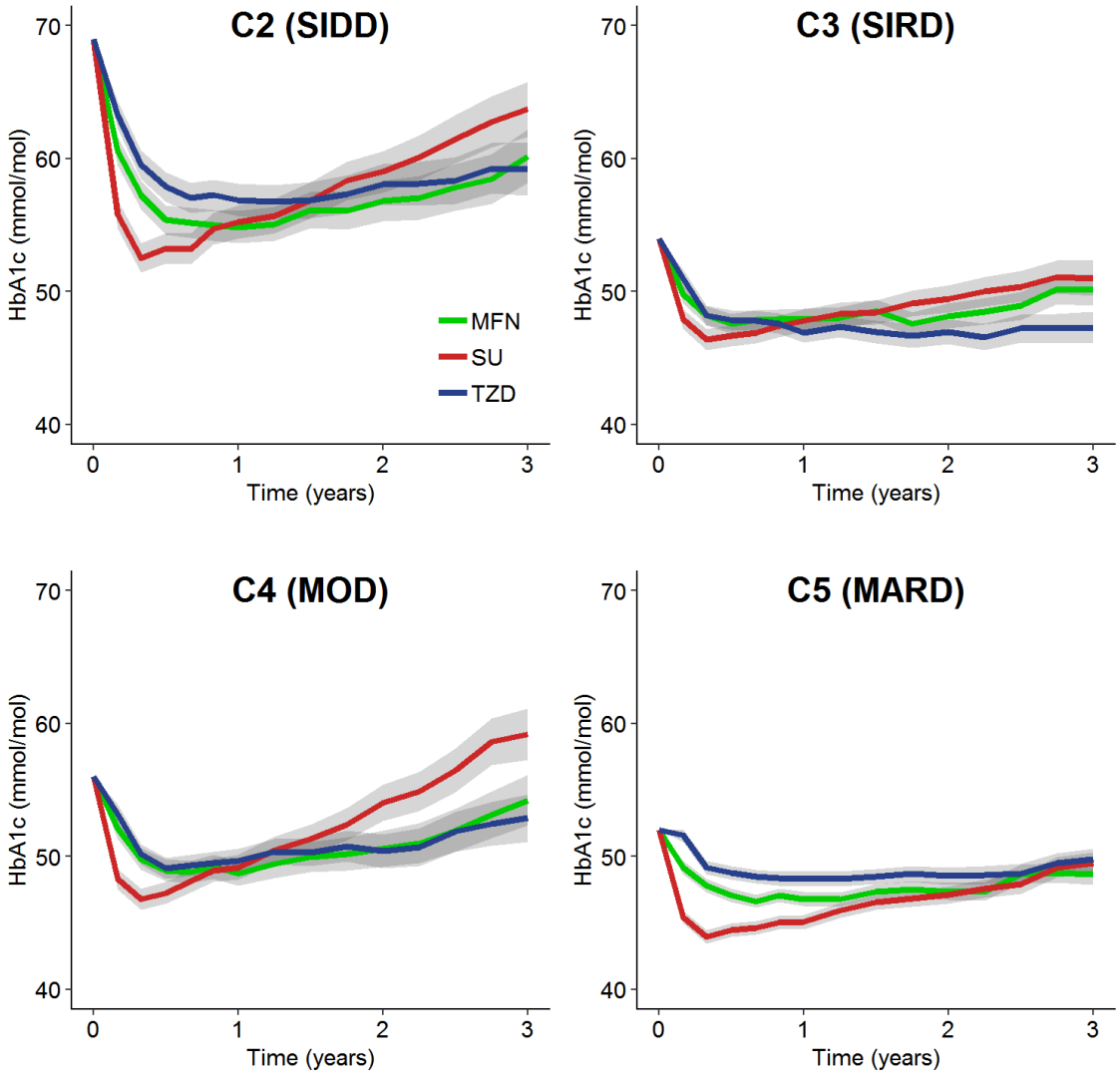


Number at risk

Cluster 1 (SAID)	129	101	91	80	65	27
Cluster 2 (SIDD)	642	499	417	325	241	94
Cluster 3 (SIRD)	637	517	429	364	300	90
Cluster 4 (MOD)	652	520	408	312	232	71
Cluster 5 (MARD)	1108	927	846	727	593	208

Patterns of HbA1c response to the different drugs differed across clusters in ADOPT (Figure 4, Supplementary Figure 8). There was an overall HbA1c benefit with thiazolidinedione therapy in cluster 3 (SIRD), and for sulfonylurea therapy in cluster 5 (MARD) (Table 2). However, the combined clinical features explained more variation in response than the clusters: R^2 was lower for Strategy A) clusters than Strategy B) clinical features (ADOPT R^2 clusters: 0.15 for metformin, 0.20 sulfonylureas, 0.17 thiazolidinediones; R^2 clinical features: 0.35 metformin, 0.33 sulfonylureas, 0.32 thiazolidinediones).

Figure 4: **Change in HbA1c by drug for clusters 2-5 in ADOPT over three years (n=3,607).** Adjusted mean HbA1c over three years by drug. Grey shading shows 95% CIs. For cluster 1 (n=158) see Supplementary Figure 8.

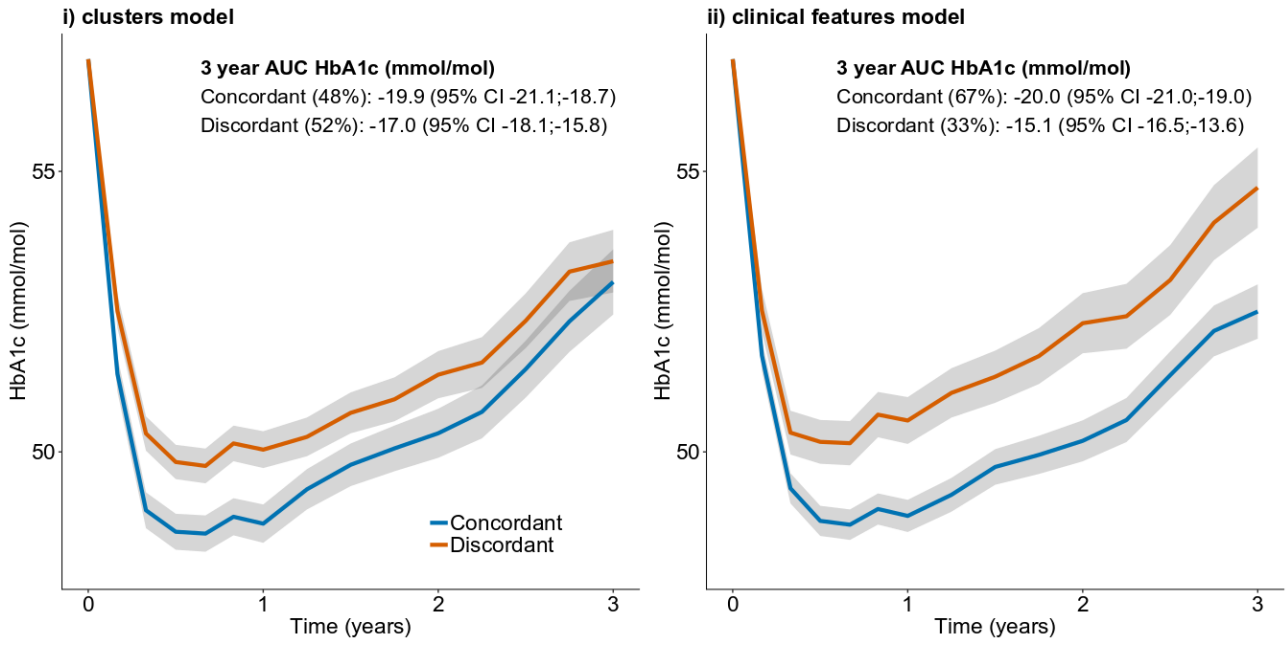


In the independent trial (RECORD) we found clinical features outperformed the clusters for treatment selection. In RECORD we tested the performance for treatment selection of the two strategies developed in ADOPT (Strategy A: selecting therapy based on predicted response to each drug at cluster level; Strategy B selecting therapy based on predicted response to each drug at the individual level based on precise clinical features (see Supplementary Table 9 for ADOPT model coefficients for the clinical features)). Each individual in each trial was assigned as concordant or discordant with the treatment selection rule under Strategy A) clusters and Strategy B) clinical features (Table 2, Supplementary Tables 10-13).

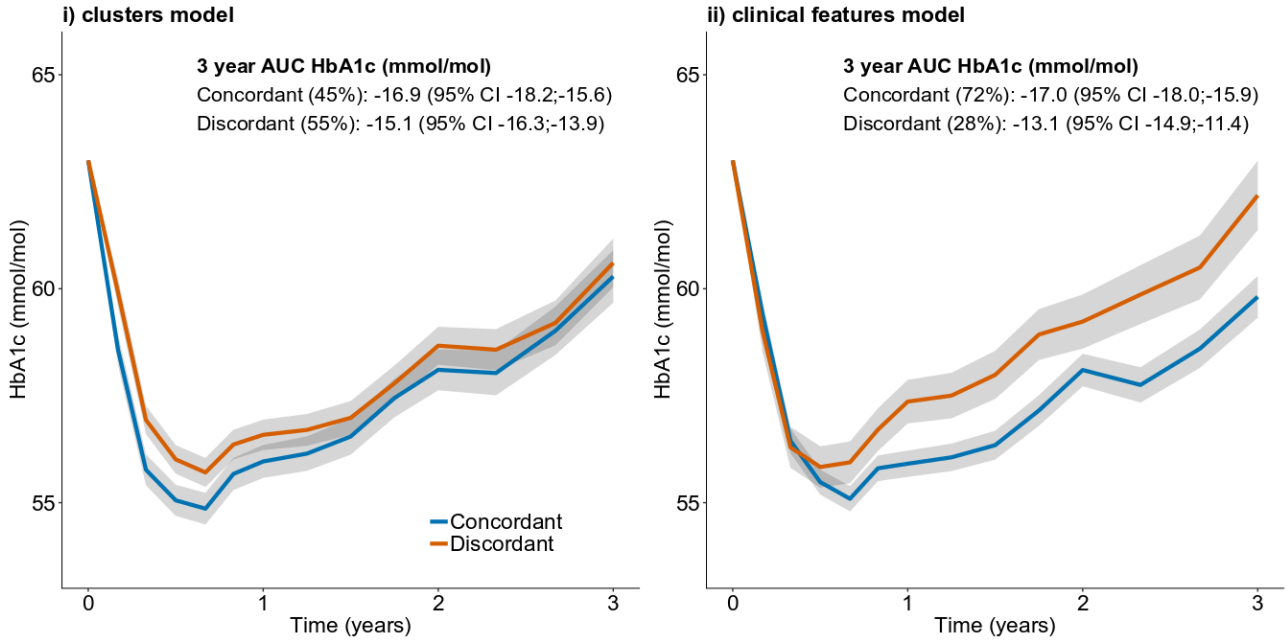
In ADOPT, with both Strategy A) Clusters and Strategy B) clinical features, there was a greater overall HbA1c reduction in the concordant group compared with the discordant group (Figure 5A). In RECORD (validation dataset) there was a greater benefit in the concordant group with Strategy B) clinical features (Figure 5B) than Strategy A) clusters. Strategy B) clinical features outperformed the clusters at all HbA1c thresholds assessed to define concordant and discordant groups in RECORD (Supplementary Table 14).

Figure 5: **Change in HbA1c over three years in concordant and discordant treatment selection groups for i) clusters model and ii) clinical features model**

(A) ADOPT development cohort (n=3,785)



(B) RECORD validation cohort (n=4,057)



There was no evidence of differences between clusters in the risk of cardiovascular hospitalisation or death in RECORD after adjustment for age (Supplementary Figure 9, Supplementary Table 15).

Clusters assigned in ADOPT using the ANDIS cluster centre coordinates were broadly similar to those defined de-novo in ADOPT (Figure 1, Supplementary Figure 10). 58% of individuals were assigned to the same cluster using both the ANDIS derived clusters and de-novo ADOPT clusters (Supplementary Table 16). Differences in outcomes by ANDIS-assigned cluster are shown in the Supplementary Figures 11-13. ADOPT clusters outperformed the ANDIS clusters for treatment response; model performance measures were similar for glycaemic and renal progression (Supplementary Table 17).

Discussion

We found the data-driven clusters of Ahlqvist and colleagues were reproducible in trial data. Clusters differed in glycaemic and renal progression but simple clinical factors features (respectively, age at diagnosis and baseline renal function) performed as well or better to predict progression. To our knowledge, for the first time we show differences by cluster in treatment response. However, clusters were markedly outperformed by models using simple clinical features for both the prediction of glucose-lowering response and for treatment selection. Overall the results suggest there will be greater clinical utility from modelling clinical features directly, rather than from using clinical features to place patients into subgroups (Figure 6).

Even though there were restricted entry criteria for both the ADOPT and RECORD trials, cluster analysis defined subgroups were very similar to those seen in non-selective Scandinavian cohorts, and subsequently Chinese and US cohorts.(4, 20) This suggests that if the cluster analysis is repeated in the specified way in new datasets it will routinely produce similar clusters.

A key strength of trial data over previous observational data is the availability of protocol driven follow-up, meaning we were able to conduct a systematic assessment and demonstrate the clusters do differ in disease progression. This is a considerable advantage over the previously described routine follow-up where therapy introduction is not protocol driven.(4) Independently of therapy, clusters 1 (SAID), 2 (SIDD) and 4 (MOD) had an increased rate of glycaemic progression.

Differences in the development of renal failure had previously been shown in observational follow-up, and we replicated a faster progression of renal disease in clusters 3 (SIRD) and 5 (MARD), although there was no evidence of a difference in renal progression after accounting for baseline renal function.

We were able to establish that the clusters differ in response to different glucose-lowering therapies. This was possible due to the randomised, systematic therapy given. We found a particular benefit for cluster 3 (SIRD) with thiazolidinediones, and for cluster 5 (MARD) with sulfonylureas.

The fact that clusters are reproducible and can help predict progression and response to therapy is important. However a key question raised in response to the original article is whether it is more clinically useful to use clinical features to assign a patient to a subgroup and then treat in a way that is best for that subgroup, or to use clinical features to predict patient outcomes directly using outcome-specific models.(7) We found simple clinical features were similar or better than the clusters to stratify disease progression and to personalise therapy. A simple model incorporating just age at diagnosis was able to predict glycaemic progression as well as the clusters, having been identified as a key predictor of progression in recent observational analysis.(21) Similarly, baseline renal function explained differences between the clusters in risk of renal progression.

For treatment response we found that models combining four simple clinical measures (age, sex, baseline HbA1c and BMI) explained more variation in response than the clusters. However, this gives little insight into which of the two approaches

is more useful to select between treatment options for an individual patient.(17, 18) A more useful test in this context is to compare the population-level effect on glycaemic response of applying each approach to select treatment.(18) We were able to directly assess this, by comparing the two approaches developed in ADOPT in an independent trial dataset (RECORD). This was possible as some participants in RECORD were randomised to the drug estimated to be 'best' for them using the ADOPT models (concordant group) whilst the remainder were randomised to a not 'best' drug (discordant group). The difference in HbA1c between the two groups provided a measure of the population-level effect of each treatment selection strategy. In RECORD we found a small benefit (1.8 mmol/mol over three years) of selecting therapy by cluster; in contrast there was a greater benefit (3.9 mmol/mol) selecting treatment using the clinical features model. These results suggest that attempts to personalise treatment in type 2 diabetes will have most clinical utility if based on the use of continuous phenotypic measures, rather than subgroup assignment.

Strengths of this study include the use of data from two large, long-term, randomised trials, in which we were able to not only reproduce the clustering approach of Ahlqvist and colleagues, but to describe diabetes progression and treatment response in protocol-driven follow-up. Furthermore we were able to test treatment selection based on clusters compared to clinical features in an independent validation dataset. The treatment selection rule we applied was designed to test clinical utility in this study, rather than to maximise outcomes for the population or individuals. Approaches to evaluate treatment selection strategies are not well-developed and are the subject of on-going methodological research.(17) A limitation

of our study is the potential non-representativeness of participants due to the original trial exclusion criteria. Both ADOPT and RECORD had exclusion criteria based on blood glucose levels and age (and BMI in RECORD); these clinical variables informed the cluster analysis. Despite this we found that the clusters were reproducible, with a pattern of differences in phenotypic measures that closely matched those previously reported. Given the variables informing the cluster analysis are not independent and are likely to be similarly correlated in most patients with diabetes, this reproducibility is not surprising,(7) although similarly to the original study we lacked data on non-white ethnicities (ADOPT was 88% Caucasian, RECORD 99%). Due to the design of the trials we were unable to evaluate some outcomes explored in the original study such as time to insulin, and we lacked power to evaluate other outcomes including development of end-stage renal disease. A further limitation was the therapy used in the trials; evaluation of heterogeneity in treatment response for the newer drug classes DPP4 inhibitors, SGLT2 inhibitors and GLP-receptor agonists would be of considerable interest.

An important difference between this study and the original study by Ahlqvist and colleagues was in the analysis of renal progression. Whilst we excluded individuals with pre-existing kidney disease, in the Scandinavian population-based cohorts people with pre-existing kidney disease when diagnosed with diabetes were not excluded and the onset of renal dysfunction was set to the first time that an abnormal value was found on clinical testing post diabetes diagnosis.

Precision medicine is successfully established in monogenic and neonatal diabetes, where it has been possible to define discrete etiological subtypes with differing genetic causes that have very different optimal treatment requirements.(22-24) A key

difference from type 2 diabetes is that in these cases the subgroups have discrete and non-overlapping aetiologies and can be robustly defined by genetic sequencing. In contrast, the study of Ahlqvist and colleagues and other recent attempts to characterise the heterogeneity in type 2 diabetes have identified clusters with limited clinical utility as they are non-aetiological, overlapping, highly dependent on the variables used to classify them and cannot be robustly defined at an individual level.(4, 25) Even genetic susceptibility clusters, which do have the advantage of being fixed throughout life, have not led to the identification of discrete etiological diabetes subtypes, although they offer insight into mechanistic pathways underlying heterogeneity.(26)

The known heterogeneity in type 2 diabetes, together with the differences we have observed in clinical outcomes, raises the possibility of a practical clinical application of precision medicine in type 2 diabetes in the near future. Our study supports the suggestion that the optimal approach to tailor management based on risk of progression and therapeutic response will be to use ‘precise’ continuous phenotypic measures to predict specific outcomes for individuals using multivariable models, rather than define subgroups and assume all individuals are homogenous within each subgroup.(7) In particular, individual clinical characteristics have been shown to have robust associations with response to specific type 2 diabetes drug options.(6, 27-29) These studies raise the possibility that the relative glucose-lowering benefit of the different drugs might be identifiable by combining simple clinical measures in a model for treatment selection. This will require systematic assessment of associations between other patient features (including lifestyle factors, biomarkers and concomitant medications) beyond those assessed in this study. The advantage

of such an approach is that the clinical features used are already part of routine clinical care. Similarly, further systematic assessment of associations between clinical patient features and glycaemic and renal progression will be required to determine whether individuals at high or low risk of progression can be robustly identified.

The methodology we have applied in this study, harnessing existing individual-level trial data to test a precision medicine strategy developed in other data, offers an exciting, low-cost framework to evaluate novel precision medicine approaches without a prospective trial. Such trial datasets are increasingly available to researchers to answer secondary research questions.(30) The approach we used of a direct comparison of different approaches in an independent data set is a good model for defining their relative performance. When defining utility of models in future studies it will be important to interrogate multiple relevant outcomes as well as glycaemia, including cardiovascular outcomes, microvascular complications, and non-glycaemic effects of specific drugs including weight change and side-effects

In conclusion, we have shown cluster-defined subgroups are reproducible and can help to define individuals that differ in the risk of diabetes progression and in glycaemic response to common therapeutic options. Our study demonstrates a ‘prediction model’ approach combining phenotypic measures to predict specific outcomes for individual patients is likely to have greater clinical utility than subgroup assignment. Existing trial data offer an exciting opportunity to evaluate the potential of precision medicine approaches to improve patient outcomes in type 2 diabetes.

References

1. Merino J, Florez JC. Precision medicine in diabetes: an opportunity for clinical translation. Annals of the New York Academy of Sciences. 2018;1411(1):140-52.
2. Gloyn AL, Drucker DJ. Precision medicine in the management of type 2 diabetes. The Lancet Diabetes & Endocrinology. 2018;6(11):891-900.
3. McCarthy MI. Painting a new picture of personalised medicine for diabetes. Diabetologia. 2017;60(5):793-9.
4. Ahlqvist E, Storm P, Käräjämäki A, Martinell M, Dorkhan M, Carlsson A, et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. The Lancet Diabetes & Endocrinology. 2018;6(5):361-9.
5. Feher MD, Munro N, Russell-Jones D, de Lusignan S, Khunti K. Novel diabetes subgroups. The Lancet Diabetes & Endocrinology. 2018;6(6):439.
6. Dennis JM, Henley WE, Weedon MN, Lonergan M, Rodgers LR, Jones AG, et al. Sex and BMI Alter the Benefits and Risks of Sulfonylureas and Thiazolidinediones in Type 2 Diabetes: A Framework for Evaluating Stratification Using Routine Clinical and Individual Trial Data. Diabetes Care. 2018;41(9):1844-53.
7. van Smeden M, Harrell FE, Jr., Dahly DL. Novel diabetes subgroups. The Lancet Diabetes & Endocrinology. 2018;6(6):439-40.
8. Kahn SE, Haffner SM, Heise MA, Herman WH, Holman RR, Jones NP, et al. Glycemic durability of rosiglitazone, metformin, or glyburide monotherapy. The New England journal of medicine. 2006;355(23):2427-43.
9. Home PD, Pocock SJ, Beck-Nielsen H, Curtis PS, Gomis R, Hanefeld M, et al. Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes (RECORD): a multicentre, randomised, open-label trial. Lancet (London, England). 2009;373(9681):2125-35.
10. Levy JC, Matthews DR, Hermans MP. Correct homeostasis model assessment (HOMA) evaluation uses the computer program. Diabetes Care. 1998;21(12):2191-2.
11. Zinman B, Kahn SE, Haffner SM, O'Neill MC, Heise MA, Freed MI. Phenotypic Characteristics of GAD Antibody-Positive Recently Diagnosed Patients With Type 2 Diabetes in North America and Europe. Diabetes. 2004;53(12):3193-200.
12. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, 3rd, Feldman HI, et al. A new equation to estimate glomerular filtration rate. Annals of internal medicine. 2009;150(9):604-12.
13. Stratton IM, Adler AI, Neil HAW, Matthews DR, Manley SE, Cull CA, et al. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. BMJ. 2000;321(7258):405-12.
14. Hennig C. Cluster-wise assessment of cluster stability. Computational Statistics & Data Analysis. 2007;52(1):258-71.
15. Nakagawa S, Schielzeth H. A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods in Ecology and Evolution. 2013;4(2):133-42.
16. Frank E. Harrell J. Regression Modeling Strategies: Springer-Verlag; 2006.
17. Bossuyt PM, Parvin T. Evaluating Biomarkers for Guiding Treatment Decisions. Ejifcc. 2015;26(1):63-70.
18. Janes H, Pepe MS, Bossuyt PM, Barlow WE. Measuring the performance of markers for guiding treatment decisions. Annals of internal medicine. 2011;154(4):253-9.
19. European Medicines Agency. Guideline on clinical investigation of medicinal products in the treatment or prevention of diabetes mellitus CPMP/EWP/1080/00 Rev. 1. Published 14 May 2012. Available from: [https://www.ema.europa.eu/].
20. Zou X, Zhou X, Zhu Z, Ji L. Novel subgroups of patients with adult-onset diabetes in Chinese and US populations. The Lancet Diabetes & Endocrinology. 2019;7(1):9-11.
21. Donnelly LA, Zhou K, Doney ASF, Jennison C, Franks PW, Pearson ER. Rates of glycaemic deterioration in a real-world population with type 2 diabetes. Diabetologia. 2018;61(3):607-15.

22. Hattersley AT, Patel KA. Precision diabetes: learning from monogenic diabetes. Diabetologia. 2017;60(5):769-77.
23. Pearson ER, Starkey BJ, Powell RJ, Gribble FM, Clark PM, Hattersley AT. Genetic cause of hyperglycaemia and response to treatment in diabetes. Lancet (London, England). 2003;362(9392):1275-81.
24. Pearson ER, Flechtner I, Njolstad PR, Malecki MT, Flanagan SE, Larkin B, et al. Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations. The New England journal of medicine. 2006;355(5):467-77.
25. Li L, Cheng W-Y, Glicksberg BS, Gottesman O, Tamler R, Chen R, et al. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. Science translational medicine. 2015;7(311):311ra174-311ra174.
26. Udler MS, Kim J, von Grotthuss M, Bonàs-Guarch S, Cole JB, Chiou J, et al. Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis. PLoS medicine. 2018;15(9):e1002654-e.
27. Jones AG, McDonald TJ, Shields BM, Hill AV, Hyde CJ, Knight BA, et al. Markers of Beta-Cell Failure Predict Poor Glycemic Response to GLP-1 Receptor Agonist Therapy in Type 2 Diabetes. Diabetes Care. 2016;39(2):250-7.
28. DeFronzo RA, Ferrannini E, Schernthaner G, Hantel S, Elsasser U, Lee C, et al. Slope of change in HbA1c from baseline with empagliflozin compared with sitagliptin or glimepiride in patients with type 2 diabetes. Endocrinology, Diabetes & Metabolism. 2018;1(2):e00016.
29. Dennis JM, Shields BM, Hill AV, Knight BA, McDonald TJ, Rodgers LR, et al. Precision Medicine in Type 2 Diabetes: Clinical Markers of Insulin Resistance Are Associated With Altered Short- and Long-Term Glycemic Response to DPP-4 Inhibitor Therapy. Diabetes Care. 2018.
30. Strom BL, Buyse ME, Hughes J, Knoppers BM. Data Sharing — Is the Juice Worth the Squeeze? New England Journal of Medicine. 2016;375(17):1608-9.

Acknowledgements

This work was supported by the Medical Research Council (UK) (MR/N00633X/1). ATH is an NIHR Senior Investigator and a Wellcome Trust Senior Investigator (098395/Z/12/Z). AGJ is supported by an NIHR Clinician Scientist award (CS-2015-15-018). ATH, BMS and JMD are supported by the NIHR Exeter Clinical Research Facility.

Supplementary Material

Cluster assignment and characteristics (ADOPT and RECORD)

Supplementary Table 1: ADOPT cluster distributions, overall and by sex (n=4,003). SAID=severe autoimmune diabetes. SIDD=severe insulin-deficient diabetes. SIRD=severe insulin-resistant diabetes. MOD=mild obesity-related diabetes. MARD=mild age-related diabetes.

	Male		Female		Overall	
Cluster	**N**	**%**	**N**	**%**	**N**	**%**
1 SAID	94	4%	74	4%	168	4%
2 SIDD	506	22%	302	18%	808	20%
3 SIRD	448	19%	369	22%	817	20%
4 MOD	411	18%	447	26%	858	21%
5 MARD	844	37%	508	30%	1352	34%

Supplementary Table 2: Cluster centre coordinates in ADOPT

	Cluster	HbA1c	BMI	Age at diagnosis	HOMA2-B	HOMA2-IR
Females	C2 (SIDD)	1.357582	-0.438702	0.209430	-0.873420	-0.508708
	C3 (SIRD)	-0.207560	0.801772	-0.048181	1.168571	1.276217
	C4 (MOD)	-0.283972	0.282755	-0.956176	-0.257172	-0.274304
	C5 (MARD)	-0.406427	-0.570389	0.751853	-0.103295	-0.383230
Males	C2 SIDD)	1.146754	-0.334983	-0.300259	-0.780702	-0.448964
	C3 (SIRD)	-0.419911	0.021167	0.587122	1.132740	0.960985
	C4 (MOD)	0.102709	1.357982	-0.838457	0.480047	0.743829
	C5 (MARD)	-0.514633	-0.471697	0.276666	-0.366980	-0.603150

Supplementary Table 3: ADOPT Cluster characteristics by sex (n=4,003). SAID=severe autoimmune diabetes. SIDD=severe insulin-deficient diabetes. SIRD=severe insulin-resistant diabetes. MOD=mild obesity-related diabetes. MARD=mild age-related diabetes. HOMA2-B=homoeostatic model assessment 2 estimates of β-cell function. HOMA2-IR=homoeostatic model assessment 2 estimates of insulin resistance.

A) Females

	Number of participants (%)	HbA1c (mmol/mol)		BMI kg/m2		Age at diagnosis (years)		HOMA2-B (%)		HOMA2-IR (%)	
Cluster		Median	IQR	Median	IQR	Median	IQR	Median	IQR	Median	IQR
1 SAID	74 (4%)	59	50-65	33	28-38	59	51-64	63	46-87	2.4	1.5-3.4
2 SIDD	302 (18%)	69	65-75	30	27-34	57	52-64	49	38-59	2.2	1.7-2.8
3 SIRD	369 (22%)	55	49-61	39	35-43	55	49-62	102	87-125	4.3	3.8-5.0
4 MOD	447 (26%)	54	50-58	35	31-40	46	41-50	67	54-79	2.6	2.1-3.1
5 MARD	508 (30%)	53	49-57	30	27-33	64	58-68	70	58-83	2.5	1.9-3.0

B) Males

	Number of participants (%)	HbA1c (mmol/mol)		BMI kg/m2		Age at diagnosis (years)		HOMA2-B (%)		HOMA2-IR (%)	
Cluster		Median	IQR	Median	IQR	Median	IQR	Median	IQR	Median	IQR
1 SAID	94 (4%)	57	53-64	29	27-33	57	49-64	60	46-77	2.4	1.7-3.1
2 SIDD	506 (22%)	67	63-73	29	27-32	53	46-60	49	38-59	2.3	1.8-2.8
3 SIRD	448 (19%)	52	48-57	31	29-34	63	57-68	100	86-117	3.7	3.2-4.6
4 MOD	411 (18%)	57	52-63	37	34-42	49	41-54	83	68-98	3.5	3.0-4.3
5 MARD	844 (37%)	52	48-56	28	26-31	59	53-65	60	51-72	2.2	1.7-2.6

Supplementary Table 4: Cluster characteristics for each trial population

ADOPT (n=4,003). Median (interquartile range) unless stated.

	1 SAID	2 SIDD	3 SIRD	4 MOD	5 MARD
N. participants (%)	168 (4%)	808 (20%)	817 (20%)	858 (21%)	1352 (34%)
HbA1c (mmol/mol)	58 (52-64)	67 (64-74)	53 (48-60)	55 (51-61)	53 (49-56)
BMI (kg/m²)	30 (27-36)	29 (27-32)	34 (30-38)	36 (33-40)	29 (26-31)
Age at diagnosis (years)	57 (49-64)	55 (48-61)	59 (53-66)	47 (41-52)	61 (55-66)
HOMA2-B (%)*	61 (46-83)	49 (38-59)	101 (87-121)	74 (59-89)	64 (53-76)
HOMA2-IR*	2.4 (1.6-3.3)	2.3 (1.8-2.8)	4.0 (3.4-4.7)	3.1 (2.4-3.7)	2.3 (1.8-2.7)
Male sex (%)	94 (56%)	506 (63%)	448 (55%)	411 (48%)	844 (62%)
Ethnicity (% White)	158 (94%)	745 (92%)	804 (98%)	801 (93%)	1327 (98%)
Fasting glucose (mmol/l)	8.3 (7.6-9.3)	9.2 (8.4-10.2)	7.9 (7.2-8.7)	8.3 (7.5-9.2)	8.0 (7.4-8.6)
Fasting insulin (pmol/L)	108 (70-150)	93 (72-129)	208 (150-280)	158 (114-215)	96 (72-126)
Fasting C-peptide (nmol/L)	0.9 (0.6-1.3)	0.8 (0.7-1.0)	1.6 (1.4-1.8)	1.2 (1.0-1.4)	0.9 (0.7-1.1)
eGFR (ml/min per 1.73m²)**	93 (82-103)	98 (87-106)	90 (77-100)	104 (96-112)	93 (82-100)
eGFR <60 at baseline (%)**	4 (2%)	14 (2%)	41 (5%)	8 (1%)	44 (3%)
Albuminuria (mg/g)***	7 (4-16)	8 (4-17)	8 (4-18)	7 (4-19)	6 (4-13)
Albuminuria ≥ 30 at baseline (%)***	26 (16%)	126 (16%)	145 (18%)	154 (18%)	158 (12%)
HDL (mmol/L)	1.2 (1.1-1.5)	1.2 (1.0-1.5)	1.1 (1.0-1.3)	1.1 (1.0-1.4)	1.3 (1.1-1.5)
LDL (mmol/L)	3.0 (2.4-3.6)	3.3 (2.7-4.0	2.9 (2.4-3.6)	3.1 (2.5-3.7)	3.2 (2.6-3.8)
ALT (U/L)	21 (16-31)	22 (17-31)	26 (19-36)	26 (18-37)	21 (16-29)

*Calculated with HOMA2 calculator using fasting glucose and fasting C-peptide
** Calculated with CKD-EPI formula ***71 individuals with missing albuminuria at baseline

RECORD (n=4,148; ADOPT-defined clusters). Median (interquartile range) unless stated.

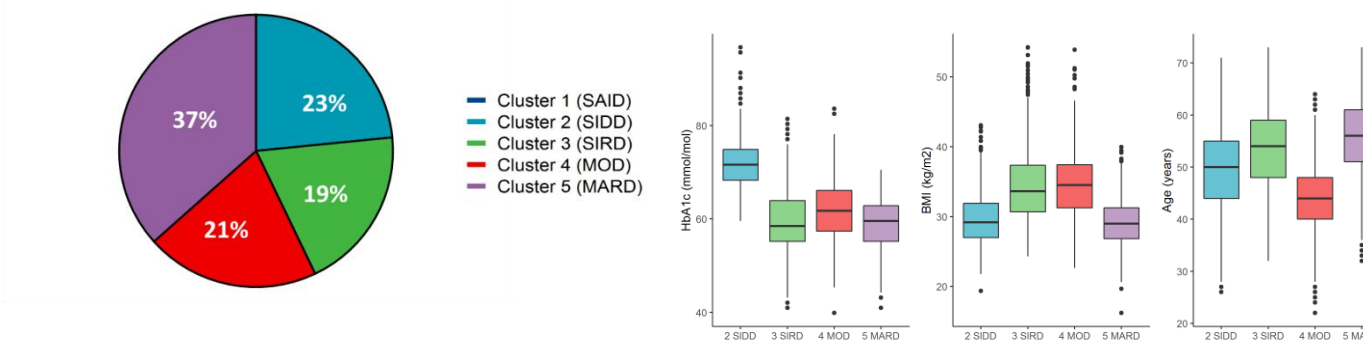
	1 SAID	2 SIDD	3 SIRD	4 MOD	5 MARD
N. participants (%)	NA	974 (23%)	803 (19%)	852 (21%)	1519 (37%)
HbA1c (mmol/mol)		72 (68-75)	58 (55-64)	62 (57-66)	60 (55-63)
BMI (kg/m²)		29 (27-32)	34 (31-37)	35 (31-37)	29 (27-31)
Age at diagnosis (years)		50 (44-55)	54 (48-59)	44 (40-48)	56 (51-61)
HOMA2-B (%)*		18 (13-24)	57 (45-74)	32 (23-42)	28 (20-36)
HOMA2-IR*		1.1 (0.7-1.5)	2.4 (1.9-3.1)	1.4 (1.0-2.0)	1.0 (0.7-1.3)
Diabetes duration (years)		7 (4-11)	5 (3-7)	6 (4-10)	5 (3-8)
Male sex (%)		571 (59%)	361 (45%)	313 (37%)	898 (59%)
Ethnicity (% White)		964 (99%)	795 (99%)	841 (99%)	1510 (99%)
Fasting glucose (mmol/l)		11 (10-13)	9 (8-10)	10 (8-11)	9 (8-10)
Fasting insulin (pmol/L)		48 (32-66)	114 (91-146)	67 (48-91)	45 (32-61)
Fasting C-peptide (nmol/L)		NA	NA	NA	NA
eGFR (ml/min per 1.73m²)**		100 (91-106)	97 (88-105)	106 (99-112)	96 (87-102)
eGFR <60 at baseline (%)**		13 (1%)	28 (3%)	9 (1%)	30 (2%)
Albuminuria (mg/g)***		9 (5-25)	9 (5-23)	9 (5-24)	8 (4-17)
Albuminuria ≥ 30 at baseline (%)***		190 (22%)	142 (20%)	149 (20%)	209 (16%)
HDL (mmol/L)		1.2 (1.0-1.4)	1.1 (0.9-1.3)	1.2 (1.0-1.4)	1.2 (1.0-1.4)
LDL (mmol/L)		3.4 (2.8-4.0)	3.2 (2.5-3.8)	3.2 (2.6-3.8)	3.3 (2.6-3.8)
ALT (U/L)		25 (19-36)	29 (21-41)	26 (19-39)	23 (17-31)

*Calculated with HOMA2 calculator using fasting glucose and fasting insulin as fasting C-peptide not available
** Calculated with CKD-EPI formula, 2 individuals missing eGFR at baseline ***479 individuals with missing albuminuria at baseline

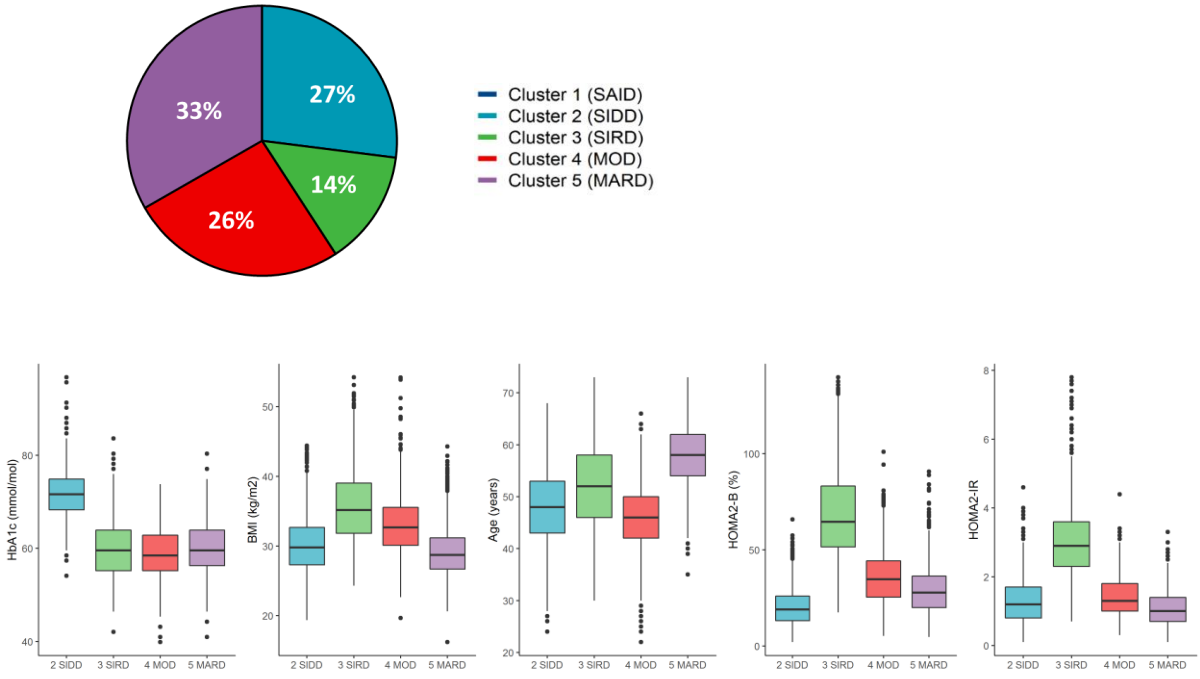
Supplementary Figure 1: clusters characteristics in RECORD. Cluster distribution and cluster characteristics (n=4,148). RECORD participants assignment and distributions of baseline clinical characteristics according to k-means clustering (A) Clusters derived in ADOPT and assigned to RECORD participants (B) Clusters derived in RECORD and assigned to RECORD participants.

SAID=severe autoimmune diabetes. SIDD=severe insulin-deficient diabetes. SIRD=severe insulin-resistant diabetes. MOD=mild obesity-related diabetes. MARD=mild age-related diabetes. HOMA2-B=homoeostatic model assessment 2 estimates of β-cell function. HOMA2-IR=homoeostatic model assessment 2 estimates of insulin resistance.

(A) ADOPT derived clusters

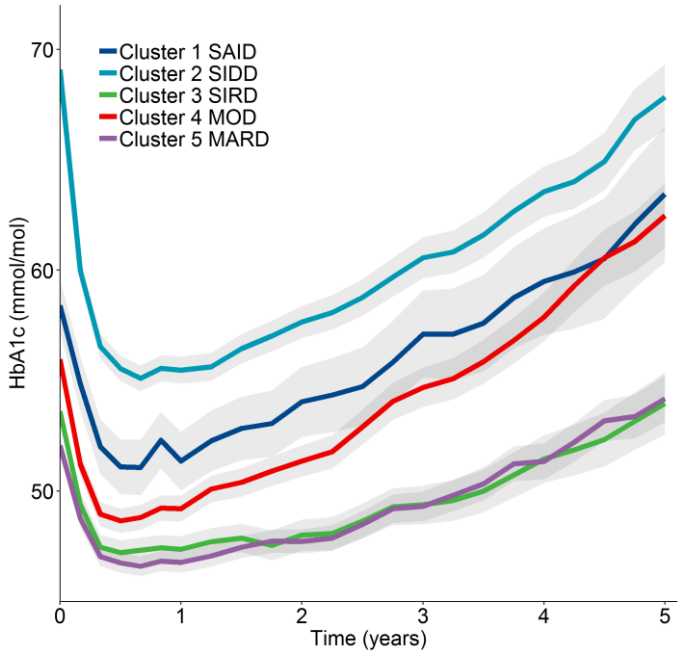


(B) RECORD derived clusters



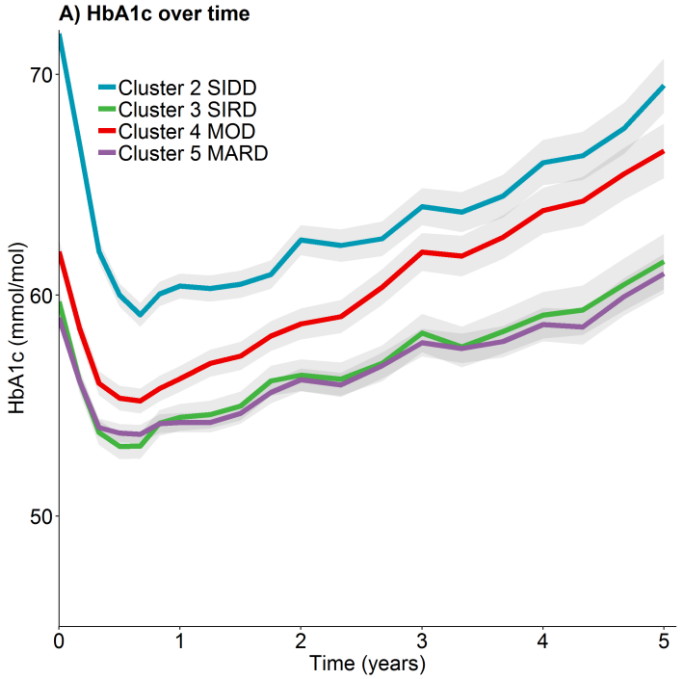
Glycaemic progression

Supplementary Figure 2: HbA1c over time from randomisation by cluster in ADOPT (n=3,802).



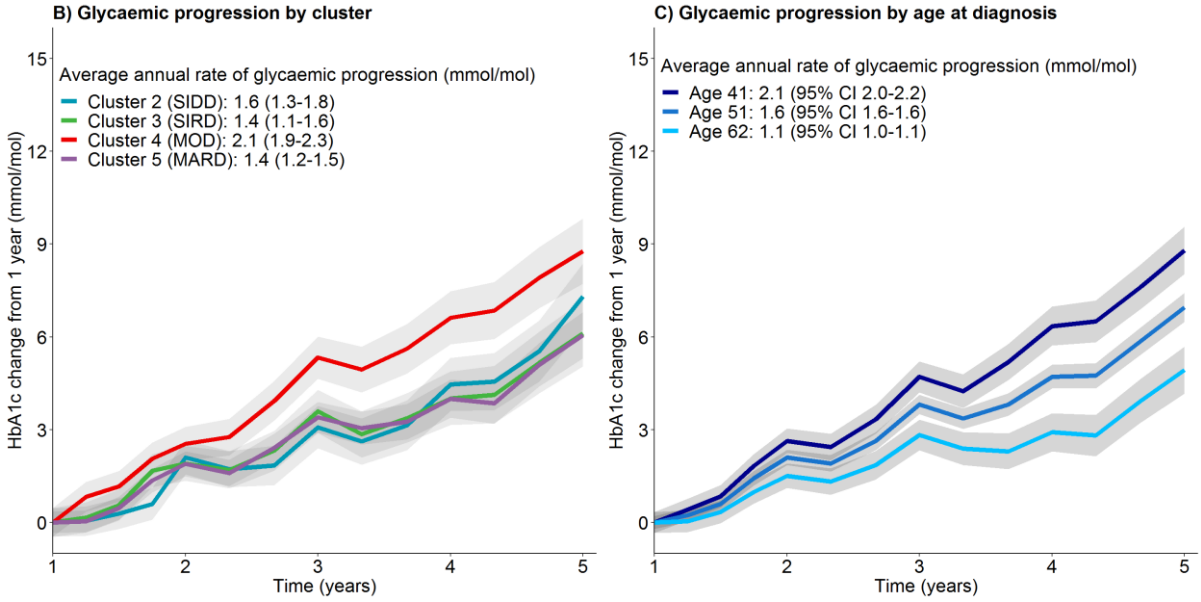
Supplementary Figure 3: HbA1c in RECORD

(A) HbA1c over time from randomisation by cluster (n=4,057);



Supplementary Figure 3 (cont.): HbA1c in RECORD.

(B) Glycaemic progression from 1 year by ADOPT derived cluster (n=3,586); (C) Glycaemic progression from 1 year by age at diagnosis (10th, 50th and 90% percentile of RECORD participants) (n=3,586). Data are estimates from repeated measures mixed effects models.



Supplementary Table 5: Glycaemic progression model performance measures to compare model using clusters and model using age at diagnosis. A higher adequacy index suggests a better model (calculated as model LR x^2 / Combined model LR x^2)

A) ADOPT

	R^2	AIC	LR x^2	Adequacy Index
Clusters	0.084	221404	1225	0.95
Age at diagnosis	0.088	221318	1210	0.94
Combined model (clusters + age at diagnosis)	0.093	221371	1292	1.00

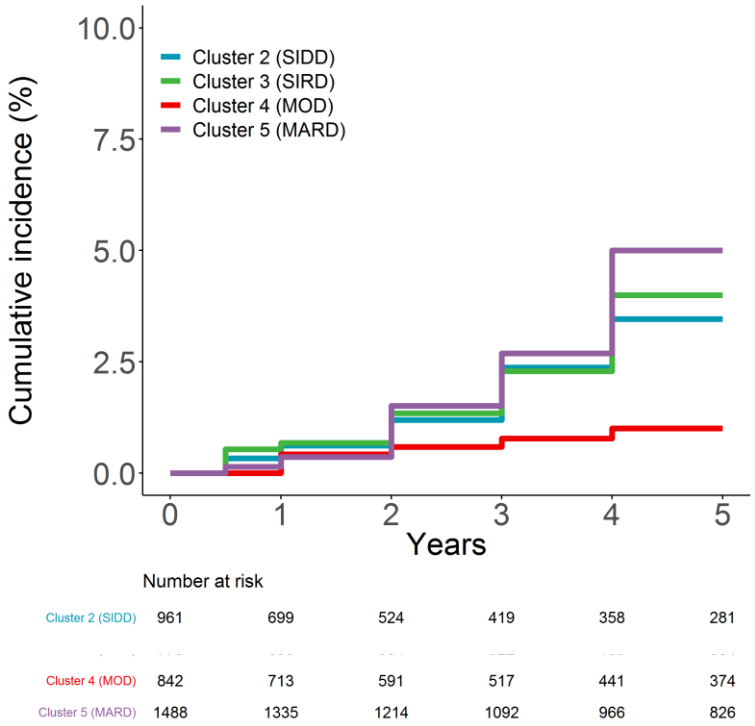
B) RECORD

	R^2	AIC	LR x^2	Adequacy Index
Clusters	0.048	274658	1065	0.89
Age at diagnosis	0.052	274624	1099	0.92
Combined model (clusters + age at diagnosis)	0.055	274642	1196	1.00

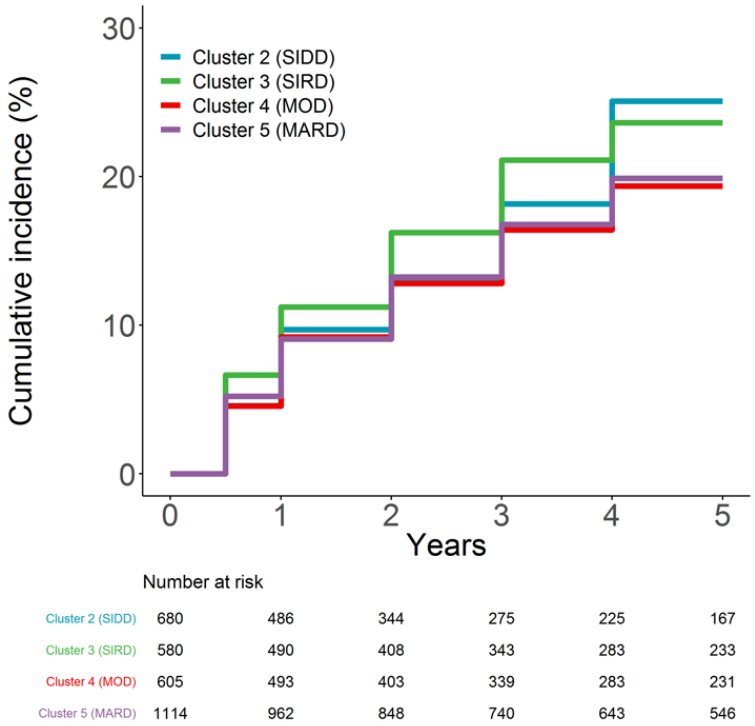
Renal progression

Supplementary Figure 4: Renal progression by cluster in RECORD (clusters derived from ADOPT)

(A) Cumulative incidence of CKD Stage 3 (confirmed eGFR <60) in individuals with eGFR ≥60 at baseline (n=4,066). eGFR calculated using CKD-EPI formula.



(B) Cumulative incidence of albuminuria (UACR ≥30 mg/g) in individuals with UACR <30 mg/g at baseline (n=2,979).



Supplementary Table 6: Risk of renal progression by cluster in RECORD (clusters derived from ADOPT)

(A) Time to CKD Stage 3 (n=4,066). eGFR calculated using CKD-EPI formula.

	No.	Person years at risk	Events	Hazard ratio (95% CI)	
				Unadjusted	Adjusted*
Time to CKD					
Cluster					
C1 (SAID)	NA	NA	NA	NA	NA
C2 (SIDD)	961	2551	17	1.00 (ref)	1.00 (ref)
C3 (SIRD)	775	2789	22	1.12 (0.60-2.11)	0.96 (0.51-1.81)
C4 (MOD)	842	2811	6	0.31 (0.12-0.78)	0.57 (0.22-1.45)
C5 (MARD)	1488	5658	55	1.37 (0.79-2.36)	1.16 (0.67-2.00)

*Adjusted for baseline eGFR

(B) Time to albuminuria (n=2,979)

	No.	Person years at risk	Events	Hazard ratio (95% CI)	
				Unadjusted	Adjusted*
Time to albuminuria					
Cluster					
C1 (SAID)	NA	NA	NA	NA	NA
C2 (SIDD)	680	1679	103	1.00 (ref)	1.00 (ref)
C3 (SIRD)	580	1860	113	1.04 (0.80-1.36)	1.02 (0.78-1.34)
C4 (MOD)	605	1869	90	0.82 (0.62-1.09)	0.82 (0.62-1.09)
C5 (MARD)	1114	3906	188	0.85 (0.66-1.08)	0.92 (0.72-1.17)

*Adjusted for baseline UACR

Supplementary Table 7: Time to CKD Stage 3. eGFR calculated using MDRD formula.

(A) ADOPT (n=3,650)

	No.	Person years at risk	Events	Hazard ratio (95% CI)	
				Unadjusted	Adjusted*
Time to CKD					
Cluster					
C1 (SAID)	152	492	7	3.00 (1.16-7.72)	1.67 (0.64-4.32)
C2 (SIDD)	748	2235	11	1.00 (ref)	1.00 (ref)
C3 (SIRD)	729	2427	35	2.99 (1.53-5.92)	1.65 (0.84-3.26)
C4 (MOD)	799	2406	11	0.93 (0.40-2.14)	1.33 (0.57-3.06)
C5 (MARD)	1222	4325	41	2.00 (1.03-3.90)	1.52 (0.78-2.97)

*Adjusted for baseline eGFR

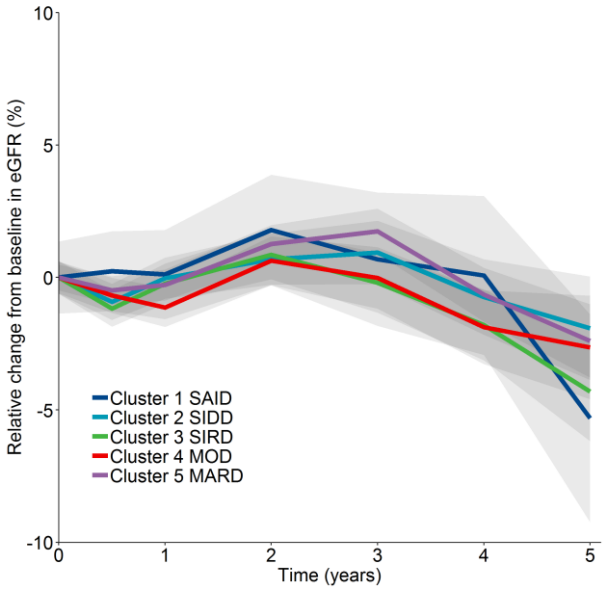
(B) RECORD (n=4,032)

	No.	Person years at risk	Events	Hazard ratio (95% CI)	
				Unadjusted	Adjusted*
Time to albuminuria					
Cluster					
C1 (SAID)	NA	NA	NA	NA	NA
C2 (SIDD)	956	2528	20	1.00 (ref)	1.00 (ref)
C3 (SIRD)	769	2753	30	1.31 (0.74-2.31)	1.10 (0.91-1.94)
C4 (MOD)	838	2781	15	0.66 (0.34-1.28)	0.98 (0.50-1.91)
C5 (MARD)	1469	5570	74	1.58 (0.96-2.59)	1.41 (0.86-2.32)

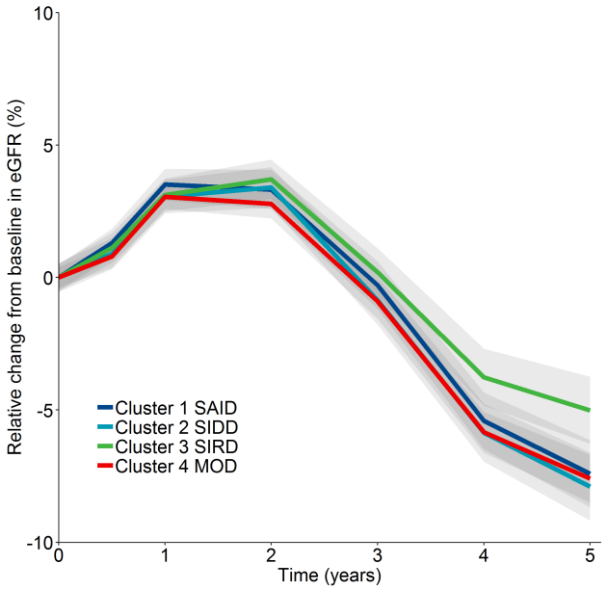
*Adjusted for baseline eGFR

Supplementary Figure 5: Relative change in eGFR from baseline, by cluster. eGFR calculated using CKD-EPI formula. Estimates are from mixed effects models.

A) ADOPT (n=3,694)

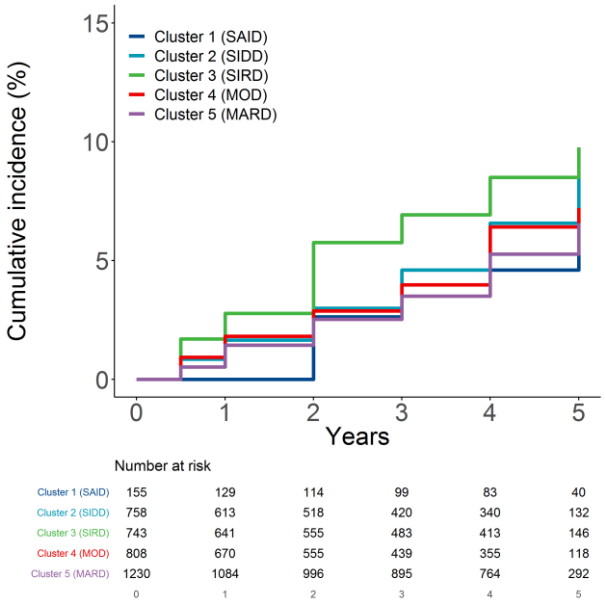


B) RECORD (n=4,066)



Supplementary Figure 6: Cumulative incidence of 30% relative change in eGFR from baseline, by cluster. eGFR calculated using CKD-EPI formula.

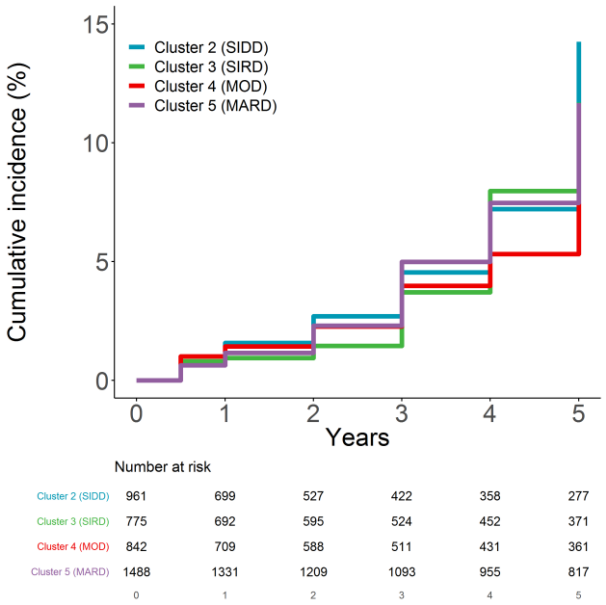
A) ADOPT (n=3,694)



Number at risk

Cluster 1 (SAID)	155	129	114	99	83	40
Cluster 2 (SIDD)	758	613	518	420	340	132
Cluster 3 (SIRD)	743	641	555	483	413	146
Cluster 4 (MOD)	808	670	555	439	355	118
Cluster 5 (MARD)	1230	1084	996	895	764	292
	0	1	2	3	4	5

B) RECORD (n=4,066)



Number at risk

Cluster 2 (SIDD)	961	699	527	422	358	277
Cluster 3 (SIRD)	775	692	595	524	452	371
Cluster 4 (MOD)	842	709	588	511	431	361
Cluster 5 (MARD)	1488	1331	1209	1093	955	817
	0	1	2	3	4	5

Supplementary Table 8: Risk of 30% relative change in eGFR from baseline by cluster. eGFR calculated using CKD-EPI formula.

(A) ADOPT (n=3,694)

	No.	Person years at risk	Events	Hazard ratio (95% CI)	
				Unadjusted	Adjusted*
Time to 30% relative change in eGFR					
Cluster					
C1 (SAID)	155	508	7	0.88 (0.39-1.97)	0.78 (0.35-1.77)
C2 (SIDD)	758	2239	35	1.00 (ref)	1.00 (ref)
C3 (SIRD)	743	2452	51	1.33 (0.87-2.05)	1.16 (0.75-1.79)
C4 (MOD)	808	2387	34	0.92 (0.57-1.48)	1.05 (0.65-1.69)
C5 (MARD)	1230	4359	54	0.79 (0.51-1.20)	0.72 (0.47-1.10)

*Adjusted for baseline eGFR

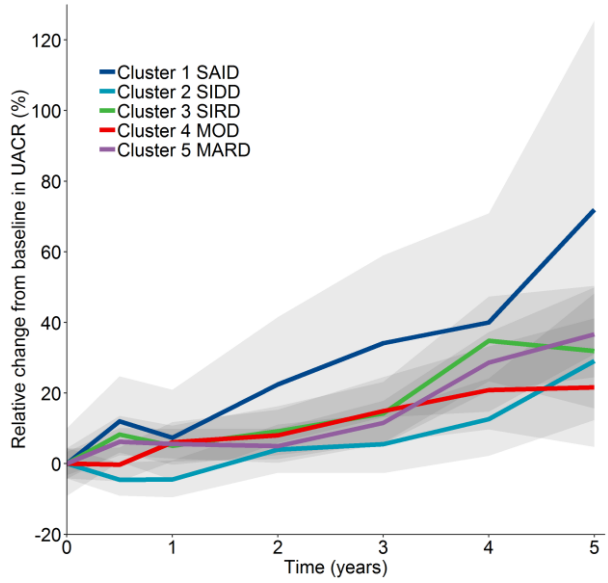
(B) RECORD (n=4,066)

	No.	Person years at risk	Events	Hazard ratio (95% CI)	
				Unadjusted	Adjusted*
Time to 30% relative change in eGFR					
Cluster					
C1 (SAID)	NA	NA	NA	NA	NA
C2 (SIDD)	961	2547	58	1.00 (ref)	1.00 (ref)
C3 (SIRD)	775	2771	57	0.83 (0.58-1.20)	0.78 (0.54-1.12)
C4 (MOD)	842	2773	40	0.59 (0.39-0.88)	0.74 (0.49-1.11)
C5 (MARD)	1488	5625	122	0.85 (0.62-1.16)	0.78 (0.57-1.07)

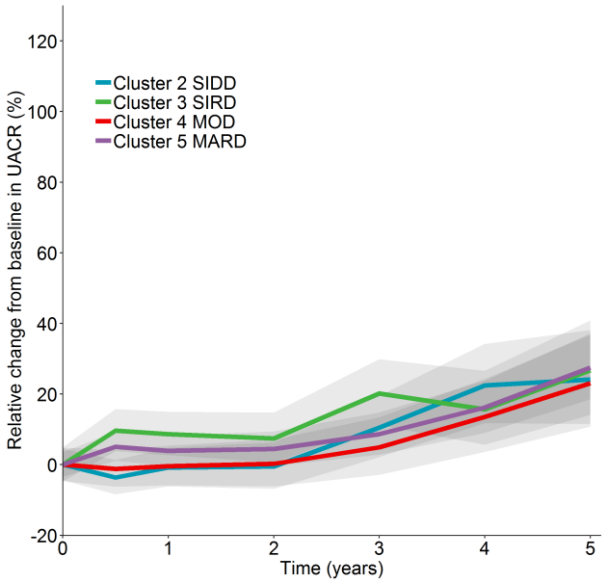
*Adjusted for baseline eGFR

Supplementary Figure 7: Relative change in urinary albumin to creatinine ratio from baseline, by cluster. Estimates are from mixed effects models with UACR modelled on log scale.

A) ADOPT (n=3,168)

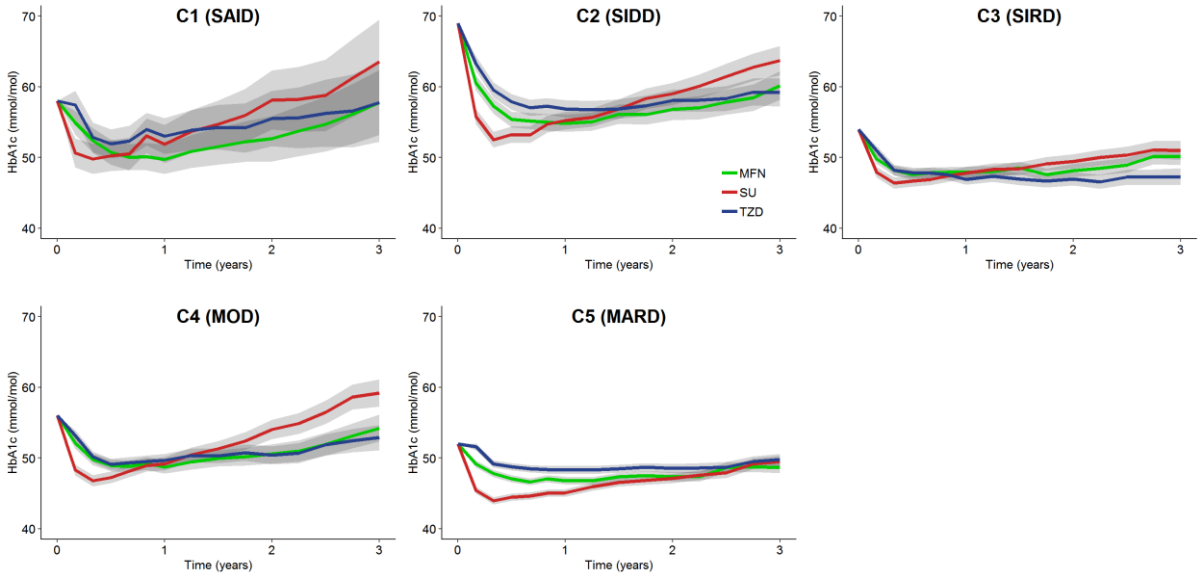


B) RECORD (n=2,979)



HbA1c response

Supplementary Figure 8: Changes in HbA1c (ADOPT trial, n=3,785). Adjusted mean HbA1c over 3 years by drug for clusters 1-5 (repeated measures mixed model). Grey shading shows 95% CIs.



Supplementary Table 9: Beta coefficients from mixed effects models for clinical features, by drug. For continuous features beta coefficients represent the change in HbA1c response for a 1-unit increase in the clinical feature. A negative coefficient indicates a higher value of the clinical feature is associated with greater reduction in HbA1c.

	Metformin	Sulfonylureas	Thiazolidinediones
Baseline HbA1c (time 0)*	0.69 (0.66;0.72)	0.59 (0.55;0.63)	0.69 (0.65;0.73)
BMI	-0.02 (-0.07;0.03)	0.03 (-0.02;0.09)	-0.11 (-0.16;-0.06)
Age at diagnosis	0.00 (-0.03;0.03)	-0.02 (-0.05;0.01)	-0.02 (-0.06;0.01)
Sex: Male	0.53 (-0.07;1.13)	-1.54 (-2.19;-0.89)	0.59 (-0.06;1.23)

*Full baseline HbA1c:study visit interaction terms not reported for brevity.

Treatment selection

Supplementary Table 10

ADOPT number of concordant individuals, by cluster, for treatment selection at 3 years based on Strategy A) treatment selection based on clusters

	Discordant	**Concordant**
Cluster		
C1 (SAID)	93	65
C2 (SIDD)	257	502
C3 (SIRD)	510	265
C4 (MOD)	272	539
C5 (MARD)	838	424

Supplementary Table 11

ADOPT number (%) of concordant individuals, by drug at 3 years, for

Strategy A) treatment selection based on clusters

	Discordant	Concordant
Overall	1970 (52%)	1795 (48%)
By randomised drug:		
Metformin	702 (55%)	569 (45%)
Sulfonylureas	555 (45%)	672 (55%)
Thiazolidinedione	713 (56%)	554 (44%)

Strategy B) treatment selection based on clinical features

	Discordant	**Concordant**
Overall	1227 (33%)	2538 (67%)
By randomised drug:		
Metformin	225 (18%)	1046 (82%)
Sulfonylureas	455 (37%)	772 (63%)
Thiazolidinedione	547 (43%)	720 (57%)

Supplementary Table 12

RECORD number of concordant individuals, by cluster, for treatment selection at 3 years based on Strategy A) treatment selection based on clusters

	Discordant	Concordant
Cluster		
C1 (SAID)	-	-
C2 (SIDD)	455	493
C3 (SIRD)	406	386
C4 (MOD)	239	594
C5 (MARD)	1121	363

Supplementary Table 13

RECORD number (%) of concordant individuals, by drug at 3 years, for a) treatment selection based on clusters

Strategy A) treatment selection based on clusters

	Discordant	Concordant
Overall	2221 (55%)	1836 (45%)
By randomised drug:		
Metformin	540 (54%)	463 (46%)
Sulfonylureas	469 (46%)	546 (54%)
Thiazolidinedione	1212 (59%)	827 (41%)

Strategy B) treatment selection based on clinical features

	Discordant	Concordant
Overall	1117 (28%)	2940 (72%)
By randomised drug:		
Metformin	23 (2%)	980 (98%)
Sulfonylureas	494 (49%)	521 (51%)
Thiazolidinedione	600 (29%)	1439 (71%)

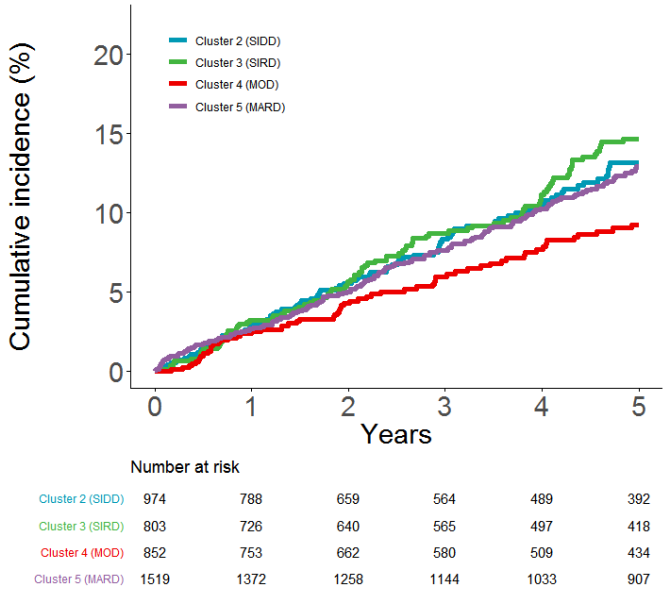
Supplementary Table 14: Cumulative HbA1c reduction at 3 years in concordant and discordant treatment selection groups using different HbA1c thresholds to define concordant/discordant groups, for clusters model and clinical features model (RECORD n=4,057)

HbA1c threshold (mmol/mol	Clusters 3 Year AUC HbA1c		Continuous features 3 Year AUC HbA1c	
	Concordant	Discordant	Concordant	Discordant
0	-18.0 (-19.6;-16.4)	-15.0 (-16.1;-14.0)	-18.3 (-20.0;-16.7)	-14.8 (-15.9;-13.8)
1	-17.0 (-18.4;-15.6)	-15.2 (-16.3;-14.0)	-18.3 (-19.6;-16.9)	-13.9 (-15.1;-12.7)
2	-17.0 (-18.4;-15.6)	-15.2 (-16.3;-14.0)	-17.6 (-18.7; -16.5)	-13.2 (-14.7;-11.8)
3	-16.9 (-18.2;-15.6)	-15.1 (-16.3;-13.9)	-17.0 (-18.0;-15.9)	-13.1 (-14.9;-11.4)
4	-16.9 (-18.1;-15.7)	-14.9 (-16.2;-13.6)	-16.6 (-17.5;-15.6)	-13.4 (-15.4;-11.4)

Cardiovascular outcomes (RECORD trial)

Supplementary Figure 9: Cumulative incidence of cardiovascular hospitalisation or death, by ADOPT-derived cluster.

RECORD (n=4,057)



Supplementary Table 15: Risk of cardiovascular hospitalisation or death by cluster in RECORD (clusters derived from ADOPT)

RECORD (n=4,057)

	No.	Person years at risk	Events	Hazard ratio (95% CI)	
				Unadjusted	Adjusted*
Time to cardiovascular hospitalisation or death					
Cluster					
C1 (SAID)	NA	NA	NA	NA	NA
C2 (SIDD)	948	3172	88	1.00 (ref)	1.00 (ref)
C3 (SIRD)	792	3038	94	1.11 (0.83-1.49)	1.06 (0.79-1.41)
C4 (MOD)	833	3141	62	0.71 (0.51-0.98)	1.02 (0.73-1.43)
C5 (MARD)	1484	5996	161	0.97 (0.74-1.25)	0.79 (0.61-1.03)

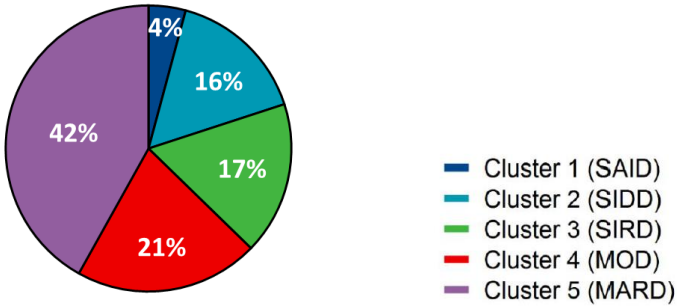
*Adjusted for age at trial entry

Application of clusters from the Swedish All New Diabetics in Scania cohort (ANDIS) to ADOPT

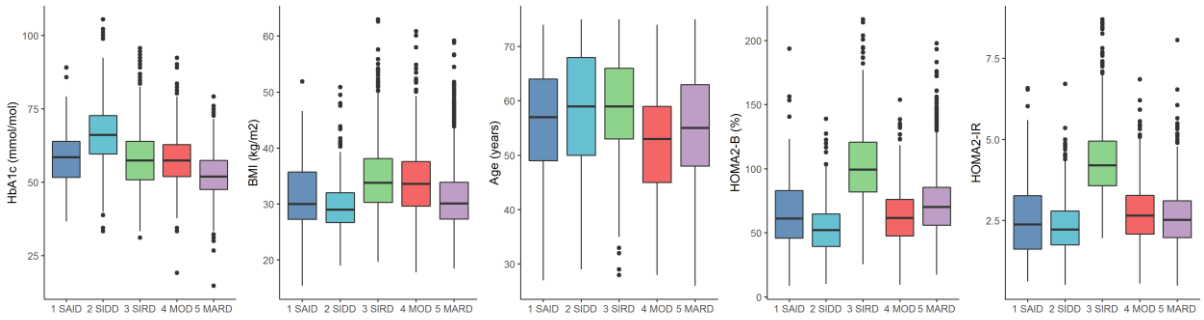
Supplementary Figure 10: Characteristics of clusters assigned in ADOPT from the cluster centre coordinates in ANDIS (n=4,003). Cluster centre coordinates originally published in Table S3, Ahlqvist et al., Lancet Diabetes Endocrinology 2018;6:361-69.

SAID=severe autoimmune diabetes. SIDD=severe insulin-deficient diabetes. SIRD=severe insulin-resistant diabetes. MOD=mild obesity-related diabetes. MARD=mild age-related diabetes. HOMA2-B=homoeostatic model assessment 2 estimates of β-cell function. HOMA2-IR=homoeostatic model assessment 2 estimates of insulin resistance.

(A) Distribution of ADOPT participants according to ANDIS clustering



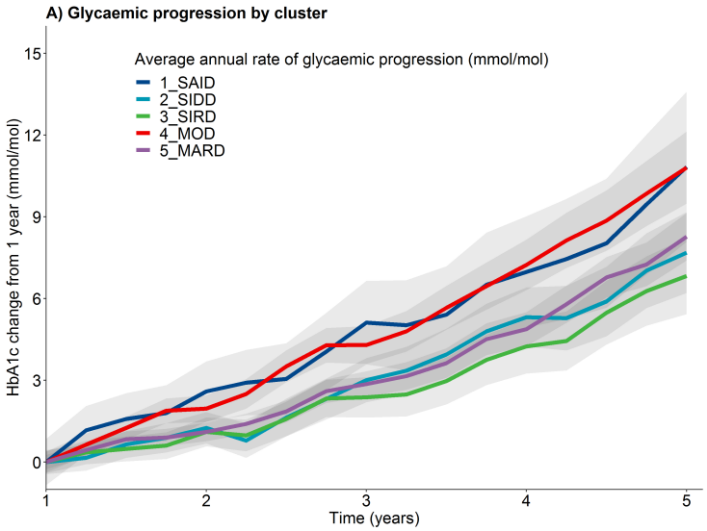
(B) Distributions of HbA1c, BMI, age at diagnosis, HOMA2-B, and HOMA2-IR at baseline for each ANDIS-derived cluster.



Supplementary Table 16: Concordance between clusters defined de-novo in ADOPT and clusters assigned in ADOPT from ANDIS cluster centre coordinates

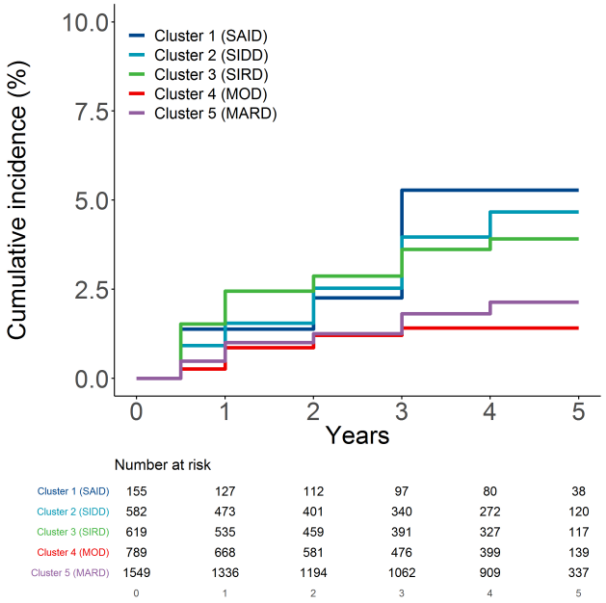
	ANDIS clusters				
ADOPT clusters	C1 (SAID)	C2 (SIDD)	C3 (SIRD)	C4 (MOD)	C5 (MARD)
C1 (SAID)	**100%**	0%	0%	0%	0%
C2 (SIDD)	0%	**56%**	9%	25%	9%
C3 (SIRD)	0%	1%	**59%**	2%	38%
C4 (MOD)	0%	2%	12%	**43%**	43%
C5 (MARD)	0%	11%	3%	18%	**68%**

Supplementary Figure 11: Glycaemic progression by cluster in ADOPT from one to five years using ANDIS-derived clusters (n=3,016)

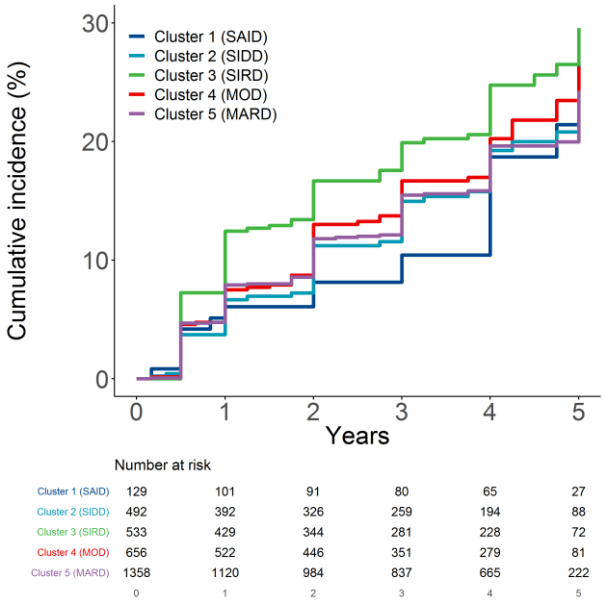


Supplementary Figure 12: Renal progression by cluster in ADOPT over five years using ANDIS-derived clusters.

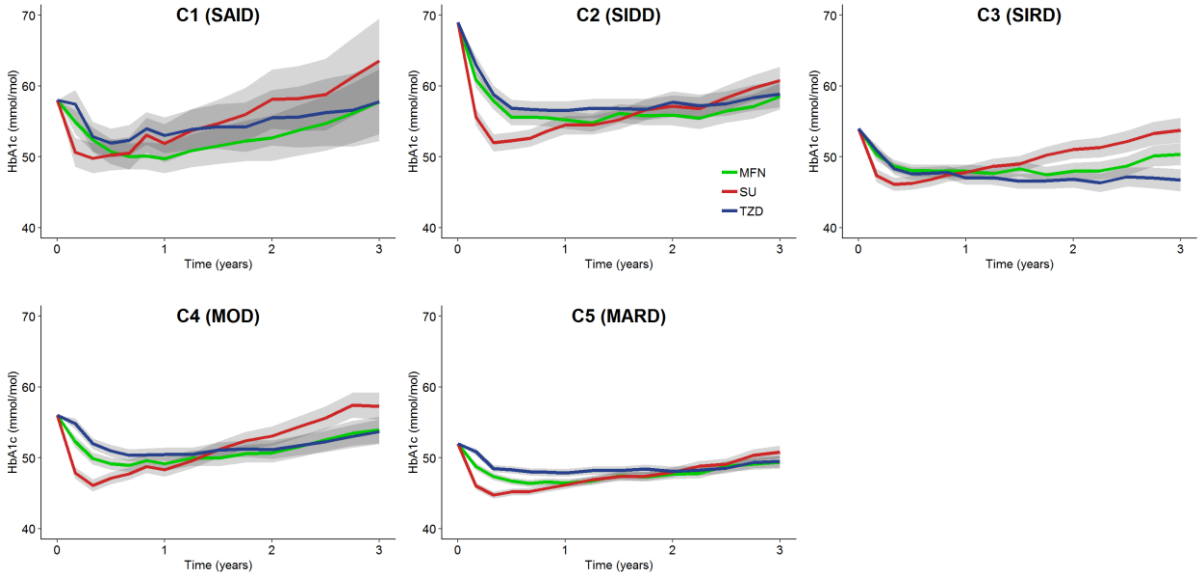
(A) Cumulative incidence of CKD Stage 3 (confirmed eGFR <60) in individuals with eGFR ≥60 at baseline (n=3,694). eGFR calculated using CKD-EPI formula.



(B) Cumulative incidence of albuminuria (UACR ≥30 mg/g) in individuals with UACR <30 mg/g at baseline (n=3,168).



Supplementary Figure 13: Change in HbA1c by drug for each cluster in ADOPT over three years using ANDIS-derived clusters (n=3,785). Adjusted mean HbA1c over three years by drug. Grey shading shows 95% CIs.



Supplementary Table 17: Model performance measures to compare clusters defined de-novo in ADOPT and clusters assigned in ADOPT from ANDIS cluster centre coordinates

A) Glycaemic progression from one to five years (n=3,016)

	R^2	AIC
ADOPT clusters	0.084	221404
ANDIS clusters	0.078	221446

B) Time to CKD Stage 3 (confirmed eGFR <60) in individuals with eGFR ≥60 at baseline (n=3,694). eGFR calculated using CKD-EPI formula.

	C-statistic	R^2
ADOPT clusters	0.58	0.01
ANDIS clusters	0.59	0.01

C) Time to albuminuria (UACR ≥30 mg/g) in individuals with UACR <30 mg/g at baseline (n=3,168).

	C-statistic	R^2
ADOPT clusters	0.52	0.002
ANDIS clusters	0.52	0.003

D) Explained variation (R^2) in treatment response (changes in HbA1c over 3 years)

	Metformin	Sulfonylurea	Thiazolidinedione
ADOPT clusters	0.15	0.20	0.17
ANDIS clusters	0.10	0.12	0.09